Trifocal tensor-based 6-DOF visual servoing

Kaixiang Zhang, François Chaumette, Jian Chen

Trifocal tensor-based 6 DOF visual servoing

The International Journal of Robotics Research
XX(X):1–19

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Kaixiang Zhang[1], François Chaumette[2] and Jian Chen[1]

Abstract
This paper proposes a trifocal tensor-based approach for 6 degrees-of-freedom visual servoing. The trifocal tensor model among the current, desired, and reference views is constructed to describe the geometric relationship of the system. More precisely, to ensure the computation consistency of trifocal tensor, a virtual reference view is introduced by exploiting the transfer relationships between the initial and desired images. Instead of resorting to explicit estimation of the camera pose, a set of visual features with satisfactory decoupling properties are constructed from the tensor elements. Based on the selected features, a visual controller is developed to regulate the camera to a desired pose, and an adaptive update law is used to compensate for the unknown distance scale factor. Furthermore, the system stability is analyzed via Lyapunov-based techniques, showing that the proposed controller can achieve almost global asymptotic stability. Both simulation and experimental results are provided to demonstrate the effectiveness and robustness of our approach under different conditions and case studies.

Keywords
Visual servoing, trifocal tensor, Lyapunov methods

1 Introduction

Visual servoing is aimed at closing the control loop with real time visual feedback to increase the flexibility, accuracy, and robustness of a robotic system (Hutchinson et al., 1996; Chaumette and Hutchinson, 2006). In this paper, the 6 degrees-of-freedom (DOF) eye-in-hand visual regulation task is considered, which is conducted via the conventional teach-by-showing idea. Specifically, an image is prerecorded in the teaching process to express a desired pose, and then by utilizing the visual feedback, the robot is driven from an initial pose to the desired pose automatically. According to the feature information used for the feedback signals, the visual servoing can be mainly divided into image-based methods (Liu et al., 2006; Kallem et al., 2007; Collewet and Marchand, 2011; Dame and Marchand, 2011; Spica et al., 2017) and pose-based methods (Wilson et al., 1996; Lippiello et al., 2007; Fujita et al., 2007).

The image-based methods define the visual features in the 2D image space. Classical image-based methods directly use image coordinates of several points or lines to construct the system errors, which would be sensitive to feature mismatching between current and desired views. To reduce the effect of image processing errors, more dense image features can be utilized, such as kernel-based (Kallem et al., 2007), photometric-based (Collewet and Marchand, 2011), and mutual information-based (Dame and Marchand, 2011) methods. However, the interaction matrices of dense visual servoing are complicated, and thus it is hard to determine the convergence region theoretically. Different from the image-based methods, the pose-based methods construct the visual features in the 3D Cartesian space. Generally, the decoupled translation and rotation information is exploited to define the system errors, which simplifies the controller design and leads to a larger convergence region compared to image-based methods. Nevertheless, to estimate the pose information, *a priori* knowledge of the target model is required.

To eliminate the requirement on *a priori* knowledge of the target model, a good choice is to design the visual control strategies based on two-view geometry, such as homography (Malis et al., 1999; Malis and Chaumette, 2000; Fang et al., 2005; Silveira and Malis, 2012; Zhang et al., 2017) and epipolar geometry (Mariottini et al., 2007; Becerra et al., 2011). More precisely, both homography-based and epipolar-based methods construct the geometric relationship between the current and desired views to facilitate the control development. This relationship is formulated by homography matrix or fundamental (essential) matrix, which can be calculated through the corresponding feature points in different views. However, both homography-based and epipolar-based methods have drawbacks. The decomposition of homography matrix requires an initial guess of the normal vector of the scene to determine the unique solution, while the epipolar geometry becomes ill-conditioned with short baseline and with planar scenes (López-Nicolás et al., 2010). Different from homography and epipolar geometry, trifocal tensor encapsulates the intrinsic geometric correlation among three views and is independent of the observed scene.

[1]State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, China
[2]Inria, Univ Rennes, CNRS, IRISA, Rennes, France

Corresponding author:
Jian Chen, College of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang, 310027, China.
Email: jchen@zju.edu.cn

Due to this fact, the trifocal tensor has great potential in addressing visual servoing (Andreff and Tamadazte, 2016; Chen et al., 2018). The trifocal tensor based visual servoing can be divided into 1-D methods (Becerra and Sagüés, 2013; Sabatta and Siegwart, 2013) and 2-D ones (López-Nicolás et al., 2010; Chen et al., 2017). Most of these methods focus on controlling a nonholonomic wheeled mobile robot to achieve different tasks, mainly including regulation (Becerra and Sagüés, 2013; López-Nicolás et al., 2010), path following (Sabatta and Siegwart, 2013), and trajectory tracking (Chen et al., 2017). Nevertheless, there are few results that extend the trifocal tensor to address the 6 DOF visual servoing, and Shademan and Jägersand (2010) is the only one up to our knowledge.

For the trifocal tensor based visual servoing, one of the main concerns is how to define the feature information. Some of the aforementioned works (Sabatta and Siegwart, 2013; Chen et al., 2017) decompose the pose information of the camera as the system states, while some others (López-Nicolás et al., 2010; Shademan and Jägersand, 2010) directly use tensor elements to provide visual feedback. In general, to extract the camera pose from the trifocal tensor, singular value decomposition (SVD) techniques need to be exploited. Although using the explicit pose information as feedback signals can simplify the controller design, SVD-based pose extraction is complicated and is sensitive to image noise. To avoid this problem, an alternative interesting idea is to construct the visual features directly from the trifocal tensor elements. This idea has been considered in Shademan and Jägersand (2010), which uses all tensor elements as visual features to design the control scheme. However, the visual features are redundant and highly coupled, and the corresponding interaction matrix is obtained via numerical techniques without deriving its analytical expression. Under these circumstances, it is extremely difficult to guarantee the system stability theoretically. To facilitate the stability analysis, linearizing the error system around the origin point is a common method. Nevertheless, this technique can only achieve local stability and cannot determine the convergence domain. Therefore, suitably refining the tensor elements as visual features and developing a control strategy with rigorous theoretical proof are motivated.

In this paper, a 6 DOF visual servoing strategy is proposed to regulate a camera from an initial pose to a desired pose. The scene-independent trifocal tensor is used to describe the vision model. Since the trifocal tensor might suffer from the degeneration issue, a reference view is introduced to avoid this problem. Specifically, the reference view is generated with the aid of the transfer relationships associated with the initial and desired images, which ensures that the trifocal tensor across the current, desired, and reference views can be estimated consistently. To facilitate the tensor normalization, an auxiliary tensor variable is introduced. Then, 3 elements of the trifocal tensor and 6 elements of the auxiliary tensor variable are chosen based on the geometric connotation to construct decoupling visual features. The relationship between the variations of tensor features and the system inputs is derived for the control development. By utilizing the Lyapunov methodology, an adaptive controller is developed for the visual regulation task, and the unknown distance scale factor is actively compensated by an update law. Theoretical analysis is provided to prove that the proposed control scheme is almost globally asymptotically stable, which is a strong result in the field of 6 DOF visual servoing. Moreover, the performance of the developed approach is validated by simulation and experimental results.

There are distinct differences between this paper and the existing trifocal tensor based works (Becerra and Sagüés, 2013; Sabatta and Siegwart, 2013; López-Nicolás et al., 2010; Chen et al., 2017; Shademan and Jägersand, 2010). First, this paper considers the 6 DOF eye-in-hand visual regulation, while Becerra and Sagüés (2013); Sabatta and Siegwart (2013); López-Nicolás et al. (2010); Chen et al. (2017) focus on the 3 DOF vision-based control of ground mobile robots. Second, instead of constructing the visual system with simplified 1-D trifocal tensor as Becerra and Sagüés (2013); Sabatta and Siegwart (2013), the 2-D trifocal tensor is utilized in this paper to describe the visual model, which is more suitable for the 6 DOF visual servoing. Third, in our work, the tensor features are chosen based on the geometric relationship to facilitate the control development and stability analysis, while in López-Nicolás et al. (2010), the tensor elements are selected experimentally, and in Chen et al. (2017), explicit pose information are decomposed from the trifocal tensor to accomplish the tracking task. Furthermore, as aforementioned, in Shademan and Jägersand (2010), all the trifocal tensor elements are exploited in the controller design, leading to a cumbersome error system, and no analytical expression of the interaction matrix is provided. Fourth, instead of designing the control law similar to the typical proportional controller as Shademan and Jägersand (2010), an adaptive regulation control law is designed in this paper, and the theoretical analysis of the system stability and convergence domain is presented. A preliminary version of this paper was presented in Zhang et al. (2018). Compared to Zhang et al. (2018), a generation strategy for the reference view is introduced to ensure that the trifocal tensor can be estimated stably, and more rigorous and detailed Lyapunov-based analysis is presented to show the almost global asymptotic convergence of the system errors instead of local stability result given in Zhang et al. (2018). Additionally, to evaluate the proposed approach thoroughly, different simulation and experiments are conducted. The results indicate that the constructed visual features show better decoupling characteristics compared to other visual features defined from trifocal tensor. It can also be seen that the robustness of the proposed approach is satisfactory with respect to image noise and calibration errors.

The remainder of this paper is organized as follows. Section 2 presents some preliminaries to facilitate the development of this work. In Section 3, the generation algorithm of the reference view is introduced. Then, Section 4 constructs the visual features with tensor variables. The adaptive visual controller and the stability analysis are developed in Section 5. Moreover, simulation and experimental results are provided in Section 6, and conclusions are given in Section 7.

2 Preliminaries

2.1 Problem statement and notations

As shown in Figure 1, this paper focuses on the 6 DOF eye-in-hand configuration. Specifically, $\mathcal{F}_i$, $\mathcal{F}_c$, $\mathcal{F}_d$, and $\mathcal{F}^*$ denote the initial, current, desired, and reference coordinate frames of the camera, respectively. Note that the reference coordinate frame $\mathcal{F}^*$ is introduced to facilitate the construction of the trifocal tensor model, which will be further discussed in the next section. Given the desired image captured in $\mathcal{F}_d$, the objective is to develop a trifocal tensor based controller to ensure that the current camera frame $\mathcal{F}_c$ asymptotically converges from $\mathcal{F}_i$ to $\mathcal{F}_d$, i.e.,

$$\mathcal{F}_c \to \mathcal{F}_d \quad \text{as} \quad t \to \infty.$$

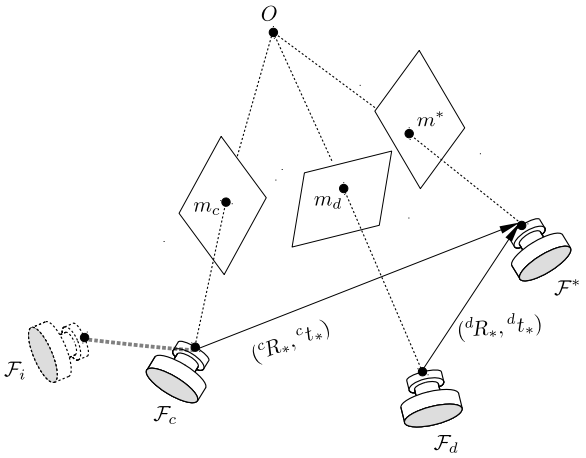

Figure 1. Trifocal tensor vision model.

To improve the readability of this paper, some notations are introduced. Throughout the paper, let $0_{n\times n}$, $I_{n\times n} \in \mathbb{R}^{n\times n}$ be the n-by-n zero and identity matrix, respectively. Let 0_n, $I_n \in \mathbb{R}^n$ be the n-by-1 vector with all zeros and ones, respectively. The subscript n might be dropped if it is clear from the context. $[\cdot]_\times \in \mathbb{R}^{3\times 3}$ is the skew symmetric matrix associated to a 3-by-1 vector, and $[\cdot]_{\times(j)}$ is the j-th column of $[\cdot]_\times$. Given a vector $c \in \mathbb{R}^n$, $c_{(j)} \in \mathbb{R}$ denotes the j-th element of c. Given a matrix $C \in \mathbb{R}^{n\times n}$, $C_{(j)} \in \mathbb{R}^n$ is the j-th column of C and $C_{(kj)} \in \mathbb{R}$ is the element on the k-th row, j-th column of C. A trifocal tensor variable $\mathcal{E} \in \mathbb{R}^{3\times3\times3}$ can be seen as a collection of three matrices $\mathcal{E}_{(1)}$, $\mathcal{E}_{(2)}$, $\mathcal{E}_{(3)} \in \mathbb{R}^{3\times3}$. Denote $\mathcal{E}_{(j)} \in \mathbb{R}^{3\times3}$ as the j-th matrix of $\mathcal{E}$. Then, $\mathcal{E}_{(jl)} \in \mathbb{R}^3$ is the l-th column of $\mathcal{E}_{(j)}$ and $\mathcal{E}_{(jkl)} \in \mathbb{R}$ is the element on the k-th row, l-th column of $\mathcal{E}_{(j)}$. Furthermore, a trifocal tensor variable, or matrix, or vector accompanied with a bracket (t) implies that its value varies with time.

2.2 Trifocal tensor geometry

The relationships between the camera frames are illustrated in Figure 1. More precisely, ${}^cR_*(t) \in \mathbb{SO}^3$ and ${}^ct_*(t) \in \mathbb{R}^3$ are the rotation and translation between $\mathcal{F}_c$ and $\mathcal{F}^*$ expressed in $\mathcal{F}_c$. Likewise, ${}^dR_* \in \mathbb{SO}^3$ and ${}^dt_* \in \mathbb{R}^3$ are the constant rotation and translation between $\mathcal{F}_d$ and $\mathcal{F}^*$ expressed in $\mathcal{F}_d$. Let $\mathcal{T}(t) \in \mathbb{R}^{3\times3\times3}$ be the trifocal tensor among the current, reference, and desired views. Then $\mathcal{T}(t)$ can be related to the pose information as follows (López-Nicolás et al., 2010; Hartley and Zisserman, 2003):

$$\mathcal{T}_{(j)} = {}^cR_{*(j)}\,{}^dt_*^T - {}^ct_*\,{}^dR_{*(j)}^T. \tag{1}$$

From (1), it is clear that the trifocal tensor encapsulates the geometric correlation across three views, and hence it is applicable for visual servoing. To estimate the trifocal tensor, point correspondences among three views are often used (Hartley and Zisserman, 2003). Consider a static feature point O in the scene, its corresponding image coordinates in the views $\mathcal{F}_c$, $\mathcal{F}_d$, and $\mathcal{F}^*$ are denoted as $p_c(t)$, p_d, $p^* \in \mathbb{R}^3$, respectively. After extracting these image coordinates from different views, the normalized Cartesian coordinates $m_c(t)$, m_d, $m^* \in \mathbb{R}^3$ can be calculated by

$$m_c = K^{-1}p_c, \qquad m_d = K^{-1}p_d, \qquad m^* = K^{-1}p^* \tag{2}$$

where $K \in \mathbb{R}^{3\times3}$ is the intrinsic camera calibration matrix. By using the point sets $(m_c(t), m_d, m^*)$, the trifocal tensor $\mathcal{T}(t)$ can be estimated up to a scale based on the following relationship (Hartley and Zisserman, 2003):

$$[m_c]_\times \left(\sum_{j=1}^{3} m^*_{(j)}\mathcal{T}_{(j)}\right)[m_d]_\times = 0_{3\times3}. \tag{3}$$

It can be seen from (1) that the camera pose can be extracted from the trifocal tensor with SVD (Hartley and Zisserman, 2003), and then this explicit pose information can be utilized easily for the controller design. Nevertheless, the SVD-based pose extraction is complex and sensitive to image noise. Hence, this work focuses on using tensor elements to construct the visual features. Inspired by Shademan and Jägersand (2010), an intuitive idea is to define the feature information with all the elements of $\mathcal{T}(t)$. In the following, the visual servoing with all tensor elements is introduced and its corresponding disadvantages are discussed.

2.3 Visual servoing with all tensor elements

Denote $s(t) \in \mathbb{R}^{27}$ as the visual features constructed with all the elements of $\mathcal{T}(t)$ as follows:

$$s \triangleq \begin{bmatrix}\mathcal{T}_{(11)}^T & \mathcal{T}_{(12)}^T & \mathcal{T}_{(13)}^T & \cdots & \mathcal{T}_{(31)}^T & \mathcal{T}_{(32)}^T & \mathcal{T}_{(33)}^T\end{bmatrix}^T. \tag{4}$$

To facilitate the control development, the time derivative of $s(t)$ should be derived. More precisely, based on the motion dynamics model, the following expression can be obtained (Chaumette and Hutchinson, 2006):

$$\begin{aligned} {}^c\dot{R}_* &= -[\omega]_\times\,{}^cR_*, & {}^c\dot{t}_* &= -v + [{}^ct_*]_\times\,\omega, \\ {}^d\dot{R}_* &= 0_{3\times3}, & {}^d\dot{t}_* &= 0_3 \end{aligned} \tag{5}$$

where $v(t), \omega(t) \in \mathbb{R}^3$ are the linear and angular velocities of the camera, respectively. Using (1), (5), and $[{}^ct_*(t)]_\times\,\omega(t) = -[\omega(t)]_\times\,{}^ct_*(t)$, the time derivative of $\mathcal{T}(t)$ can be determined, as follows:

$$\begin{aligned} \dot{\mathcal{T}}_{(j)} &= {}^c\dot{R}_{*(j)}\,{}^dt_*^T - {}^c\dot{t}_*\,{}^dR_{*(j)}^T \\ &= v\,{}^dR_{*(j)}^T - [\omega]_\times\,\mathcal{T}_{(j)} \end{aligned} \tag{6}$$

indicating that

$$\dot{\mathcal{T}}_{(jl)} = \underbrace{\begin{bmatrix} {}^{d}R_{*(lj)} I_{3\times 3} & [\mathcal{T}_{(jl)}]_{\times} \end{bmatrix}}_{L_{jl}} \begin{bmatrix} v \\ \omega \end{bmatrix}. \tag{7}$$

Taking the time derivative of (4) and utilizing (7), it can be derived that

$$\dot{s} = L \begin{bmatrix} v \\ \omega \end{bmatrix} \tag{8}$$

with $L(t) \in \mathbb{R}^{27\times 6}$ being the interaction matrix constructed by

$$L = \begin{bmatrix} L_{11}^T & L_{12}^T & L_{13}^T & \cdots & L_{31}^T & L_{32}^T & L_{33}^T \end{bmatrix}^T. \tag{9}$$

Note that $L(t)$ depends on the constant rotation ${}^{d}R_{*}$, thus an estimation or an approximation of ${}^{d}R_{*}$ will be needed in the control scheme. Based on (8), the control inputs can be designed via a typical way, as follows:

$$\begin{bmatrix} v \\ \omega \end{bmatrix} = -k\hat{L}^{+}(s - s_d) \tag{10}$$

where $k \in \mathbb{R}$ is the control gain, $\hat{L}^{+}(t) \in \mathbb{R}^{6\times 27}$ is an approximation of the pseudo-inverse of $L(t)$, and $s_d \in \mathbb{R}^{27}$ is the corresponding desired value of $s(t)$.

The aforementioned approach based on all tensor elements has several disadvantages.

• The computation of trifocal tensor might suffer from the degeneration problem during the control procedure. Consider the vision model presented in Figure 1, if the reference frame $\mathcal{F}^*$ coincides with the frame $\mathcal{F}_d$ or $\mathcal{F}_c$, then the calculation of trifocal tensor will be degenerated by using (3) (see Appendix A). Existing works (Shademan and Jägersand, 2010; Zhang et al., 2018) usually use the current, desired, and initial views to define the trifocal tensor. However, under this configuration, the aforementioned degeneration case occurs easily, typically at the beginning of the servo where the current view is coincident with the initial one, and hence it is numerically troublesome to estimate the trifocal tensor (Becerra et al., 2014). To ensure that the trifocal tensor can be calculated consistently and stably, a generation strategy for the reference view $\mathcal{F}^*$ is introduced in the next section.

• From (1), it can be found that the time-varying pose signals ${}^{c}R_{*}(t)$ and ${}^{c}t_{*}(t)$ are coupled into the expression of $\mathcal{T}(t)$. Therefore, if the tensor elements are directly chosen as the visual features, the derived interaction matrix does not present satisfactory decoupling characteristics as shown in (7). It should also be pointed out that $\mathcal{T}(t)$ can only be estimated up to a scale, i.e.,

$$\mathcal{T}_{\lambda} = \lambda \mathcal{T} \tag{11}$$

where $\mathcal{T}_{\lambda}(t) \in \mathbb{R}^{3\times 3\times 3}$ is the obtained scaled trifocal tensor and $\lambda \in \mathbb{R}$ is an unknown scale parameter. Since λ is different each time the tensor variable is estimated, to make the trifocal tensor-based approach applicable, finding a suitable way to normalize $\mathcal{T}_{\lambda}(t)$ is necessary (López-Nicolás et al., 2010). Motivated by the above issues, in Section 4, an auxiliary tensor variable is introduced to redefine the trifocal tensor-based visual features with satisfactory decoupling properties, and a normalization method is introduced to ensure that the tensor variables are scaled by a common factor during the control procedure.

• The controller designed in (10) relies on all the tensor elements and can be regarded as classical image-based control scheme. Actually, this kind of controller can only achieve local asymptotic stability, and the resulting system may be attracted to a local minimum. Due to the complex structure of the interaction matrix $L(t)$, it is quite difficult to determine the configurations with respect to the local minimum and the size of the attraction domain (Chaumette and Hutchinson, 2006). Based on the tensor features selected in Section 4, an adaptive controller is designed in Section 5. Both the local minimum and the convergence domain are analyzed clearly.

3 Reference view generation

To ensure consistent computation of trifocal tensor, a generation strategy of the reference view $\mathcal{F}^*$ is developed in this section. The core ideas are to proactively design the trifocal tensor among the initial, desired, and reference views based on the reconstructed camera pose between the initial and desired frames, and then to determine the normalized Cartesian coordinate m^* involved in (3) with classical transfer techniques (Hartley, 1997; López-Nicolás et al., 2009; Becerra et al., 2014).

The generation strategy of the reference view $\mathcal{F}^*$ is concluded in Algorithm 1. Specifically, let $m_i \in \mathbb{R}^3$ be the normalized Cartesian coordinate of the feature point O expressed in $\mathcal{F}_i$. Define $\mathcal{T}_i \in \mathbb{R}^{3\times 3\times 3}$ as the trifocal tensor with respect to the desired frame, initial frame, and the third frame being equal to the desired frame. Then the point correspondences (m_i, m_d) extracted from the initial and desired views can be used to estimate $\mathcal{T}_i$ up to a scale via the following relationship:

$$[m_d]_{\times} \left(\sum_{j=1}^{3} m_{i(j)} \mathcal{T}_{i(j)} \right) [m_d]_{\times} = 0_{3\times 3}. \tag{12}$$

Although two of the three frames used to define $\mathcal{T}_i$ are coincident, (12) can be exploited to estimate the trifocal tensor $\mathcal{T}_i$ without degeneration (see Appendix A). Denote ${}^{d}R_i \in \mathbb{SO}^3$ and ${}^{d}t_i \in \mathbb{R}^3$ as the constant rotation and translation between $\mathcal{F}_d$ and $\mathcal{F}_i$ expressed in $\mathcal{F}_d$. The trifocal tensor $\mathcal{T}_i$ can be related to the aforementioned pose signals as follows:

$$\mathcal{T}_{i(j)} = {}^{d}R_{i(j)} {}^{d}t_i^T - {}^{d}t_i {}^{d}R_{i(j)}^T. \tag{13}$$

After estimating the trifocal tensor $\mathcal{T}_i$, the essential matrix between the initial and desired views can be derived, and then classical algorithms can be used to decompose the rotation ${}^{d}R_i$ and the scaled translation ${}^{d}\bar{t}_i$ from the obtained essential matrix (Hartley and Zisserman, 2003; Ma et al., 2003). Now, let ${}^{d}R_* \in \mathbb{SO}^3$ and ${}^{d}\bar{t}_* \in \mathbb{R}^3$ be the constant rotation and scaled translation between $\mathcal{F}_d$ and $\mathcal{F}^*$ expressed in $\mathcal{F}_d$. Since $\mathcal{F}^*$ is a virtual frame, ${}^{d}R_*$ and ${}^{d}\bar{t}_*$ can be set up proactively to describe the relationship between $\mathcal{F}_d$ and $\mathcal{F}^*$. Note that the actively constructed ${}^{d}\bar{t}_*$ can be regarded as the translation scaled by the same parameter as ${}^{d}\bar{t}_i$, and thus according to the geometric correlation, the rotation ${}^{i}R_* \in \mathbb{SO}^3$ and

the scaled translation ${}^{i}\bar{t}_{*} \in \mathbb{R}^3$ between $\mathcal{F}_i$ and $\mathcal{F}^*$ can be calculated by

$$ {}^{i}R_{*} = {}^{d}R_{i}^{T}\,{}^{d}R_{*}, \qquad {}^{i}\bar{t}_{*} = {}^{d}R_{i}^{T}\left({}^{d}\bar{t}_{*} - {}^{d}\bar{t}_{i}\right). \tag{14} $$

Furthermore, by using the measurable relative pose information, the trifocal tensor $\mathcal{T}^* \in \mathbb{R}^{3\times3\times3}$ among the initial, desired, and reference views can be computed, as follows:

$$ \mathcal{T}^*_{(j)} = {}^{i}R_{*(j)}\,{}^{d}\bar{t}_{*}^{T} - {}^{i}\bar{t}_{*}\,{}^{d}R_{*(j)}^{T}. \tag{15} $$

Using $\mathcal{T}^*$, the relationship of the point correspondences m_i, m_d, and m^* is described by

$$ [m_i]_\times \left(\sum_{j=1}^{3} m^*_{(j)}\mathcal{T}^*_{(j)}\right)[m_d]_\times = 0_{3\times3}. \tag{16} $$

It is clear from (16) that given m_i, m_d, and $\mathcal{T}^*$, the least-squares solution to m^* can be calculated, which is the well known transfer issue (Hartley, 1997; Hartley and Zisserman, 2003). After obtaining m^*, the trifocal tensor $\mathcal{T}(t)$ can be determined from (3).

Algorithm 1 Reference view generation

Input:
Initial frame, $\mathcal{F}_i$, m_i;
Desired frame, $\mathcal{F}_d$, m_d;
Output:
Reference frame, $\mathcal{F}^*$, m^*;

1: Use (m_i, m_d) to estimate $\mathcal{T}_i$ from the relationship given in (12);
2: Decompose ${}^{d}R_i$ and ${}^{d}\bar{t}_i$ from the estimated $\mathcal{T}_i$;
3: Set up ${}^{d}R_*$ and ${}^{d}\bar{t}_*$;
4: Compute ${}^{i}R_*$ and ${}^{i}\bar{t}_*$ with (14);
5: Compute $\mathcal{T}^*$ with (15);
6: Compute m^* with (16).

Since the initial and desired views are recorded before starting the control task, ${}^{d}R_i$, ${}^{d}\bar{t}_i$, and m^* can be obtained off-line from Algorithm 1. It will be shown in Section 6 that our controller is robust to coarse approximation of ${}^{d}R_i$ and ${}^{d}\bar{t}_i$. Besides, owing to the proactive design of the pose signals ${}^{d}R_*$ and ${}^{d}\bar{t}_*$, the reference view $\mathcal{F}^*$ can be constructed suitably to guarantee the consistent computation of the trifocal tensor $\mathcal{T}(t)$ among the current, desired, and reference views. For example, if ${}^{d}\bar{t}_i \neq 0$, then by selecting ${}^{d}R_* = {}^{d}R_i$ and ${}^{d}\bar{t}_* = -{}^{d}\bar{t}_i$, the reference frame $\mathcal{F}^*$ will be in front or back of the frames $\mathcal{F}_i$ and $\mathcal{F}_d$ as illustrated in Figure 2. Under this circumstance, it can efficiently avoid the degeneration case that $\mathcal{F}^*$ coincides with $\mathcal{F}_d$ or $\mathcal{F}_c$ during the control procedure, and thus the trifocal tensor $\mathcal{T}(t)$ can be estimated consistently.

4 Feature construction

After estimating the trifocal tensor $\mathcal{T}(t)$, visual features need to be designed to facilitate the control development. This section focuses on defining the feature information with constructed tensor variables.

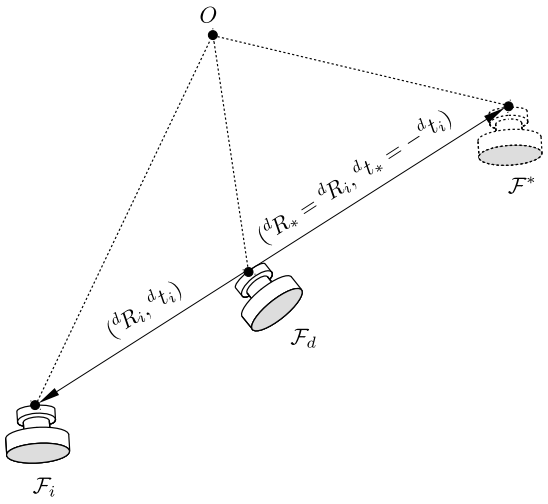

Figure 2. Geometric relationship among the initial, desired, and reference views by selecting ${}^{d}R_* = {}^{d}R_i$ and ${}^{d}\bar{t}_* = -{}^{d}\bar{t}_i$.

4.1 Auxiliary tensor variable

The expression of trifocal tensor $\mathcal{T}(t)$ given in (1) involves both the time-varying rotation ${}^{c}R_*(t)$ and translation ${}^{c}t_*(t)$. Thus, to define the visual features with decoupling characteristics, it is necessary to find a suitable way to separate ${}^{c}R_*(t)$ or ${}^{c}t_*(t)$ from $\mathcal{T}(t)$. Based on (1) and the fact that ${}^{d}R_{*(j)}^{T}\left[{}^{d}R_{*(j)}\right]_\times = 0$, this task is achieved by introducing an auxiliary tensor variable $\mathcal{Q}(t) \in \mathbb{R}^{3\times3\times3}$, as follows:

$$ \mathcal{Q}_{(j)} = \mathcal{T}_{(j)}\left[{}^{d}R_{*(j)}\right]_\times = {}^{c}R_{*(j)}\,{}^{d}t_{*}^{T}\left[{}^{d}R_{*(j)}\right]_\times. \tag{17} $$

The auxiliary tensor variable $\mathcal{Q}(t)$ is designed with the aid of the constant rotation ${}^{d}R_*$. Practically, ${}^{d}R_*$ is closely related to the trifocal tensor $\mathcal{T}(t)$ and its time derivative as shown in (1) and (7), and thus for all tensor element-based visual servoing, it would be inevitable to introduce ${}^{d}R_*$ into the feature and controller design. From (17), it can be found that for ${}^{d}t_* \neq 0_3$, $\mathcal{Q}_{(j)}(t)$ will degenerate if and only if ${}^{d}t_*$ and ${}^{d}R_{*(j)}$ are collinear. Since ${}^{d}R_{*(1)}$, ${}^{d}R_{*(2)}$, and ${}^{d}R_{*(3)}$ are linearly independent, $\mathcal{Q}_{(1)}(t)$, $\mathcal{Q}_{(2)}(t)$, and $\mathcal{Q}_{(3)}(t)$ will not suffer from degeneracy simultaneously. Furthermore, ${}^{d}R_*$ and ${}^{d}\bar{t}_*$ can be selected proactively via Algorithm 1, and thus the degeneracy of $\mathcal{Q}_{(j)}(t)$ can be avoided effectively. The corresponding estimated value of $\mathcal{Q}(t)$ is denoted as $\mathcal{Q}_\lambda(t) \in \mathbb{R}^{3\times3\times3}$. Once $\mathcal{T}_\lambda(t)$ is obtained, $\mathcal{Q}_\lambda(t)$ can be computed by

$$ \mathcal{Q}_{\lambda(j)} = \mathcal{T}_{\lambda(j)}\left[{}^{d}R_{*(j)}\right]_\times. \tag{18} $$

From (11), (17), and (18), it can be concluded that $\mathcal{Q}_\lambda(t)$ equals to $\mathcal{Q}(t)$ scaled by the unknown factor λ, i.e.,

$$ \mathcal{Q}_\lambda = \lambda\mathcal{Q}. \tag{19} $$

4.2 Tensor normalization

Since λ is different each time the tensor variables are estimated, a normalization method has to be developed to guarantee that the tensor variables are scaled by a common factor during the control procedure (López-Nicolás et al.,

2010). To facilitate the tensor normalization, a property of $\mathcal{Q}_\lambda(t)$ is presented now. According to (17) and (19), the following expression can be obtained:

$$\mathcal{Q}_{\lambda(jl)} = \lambda {}^{c}R_{*(j)} {}^{d}t_{*}^{T} \left[{}^{d}R_{*(j)} \right]_{\times(l)}. \tag{20}$$

Using (20) and the facts that ${}^{c}R_{*(j)}^{T}(t) {}^{c}R_{*(j)}(t) = 1$, ${}^{d}t_{*}^{T} \left[{}^{d}R_{*(j)} \right]_{\times(l)} = -{}^{d}R_{*(j)}^{T} \left[{}^{d}t_{*} \right]_{\times(l)}$, it can be concluded that

$$\mathcal{Q}_{\lambda(jl)}^{T} \mathcal{Q}_{\lambda(jl)} = \lambda^2 \left({}^{d}R_{*(j)}^{T} \left[{}^{d}t_{*} \right]_{\times(l)} \right)^2. \tag{21}$$

Based on (21), it can be determined that

$$\begin{aligned} \sum_{j=1}^{3} \mathcal{Q}_{\lambda(j1)}^{T} \mathcal{Q}_{\lambda(j1)} &= \lambda^2 \sum_{j=1}^{3} \left({}^{d}R_{*(2j)} {}^{d}t_{*(3)} - {}^{d}R_{*(3j)} {}^{d}t_{*(2)} \right)^2 \\ &= \lambda^2 \left({}^{d}t_{*(2)}^2 + {}^{d}t_{*(3)}^2 \right) \end{aligned} \tag{22}$$

where $\sum_{j=1}^{3} {}^{d}R_{*(2j)}^2 = \sum_{j=1}^{3} {}^{d}R_{*(3j)}^2 = 1$ and $\sum_{j=1}^{3} {}^{d}R_{*(2j)} {}^{d}R_{*(3j)} = 0$ are used. It can also be deduced that

$$\begin{aligned} \sum_{j=1}^{3} \mathcal{Q}_{\lambda(j2)}^{T} \mathcal{Q}_{\lambda(j2)} &= \lambda^2 \left({}^{d}t_{*(1)}^2 + {}^{d}t_{*(3)}^2 \right) \\ \sum_{j=1}^{3} \mathcal{Q}_{\lambda(j3)}^{T} \mathcal{Q}_{\lambda(j3)} &= \lambda^2 \left({}^{d}t_{*(1)}^2 + {}^{d}t_{*(2)}^2 \right). \end{aligned} \tag{23}$$

According to (22) and (23), the following property can be obtained:

$$\sum_{j=1}^{3} \sum_{k=1}^{3} \sum_{l=1}^{3} \mathcal{Q}_{\lambda(jkl)}^2 = 2\lambda^2 \left({}^{d}t_{*}^{T} {}^{d}t_{*} \right) = 2\lambda^2 d^{*2}. \tag{24}$$

where $d^* \triangleq \sqrt{{}^{d}t_{*}^{T} {}^{d}t_{*}} \in \mathbb{R}$ is the constant distance between $\mathcal{F}_d$ and $\mathcal{F}^*$. Since the translation between $\mathcal{F}_d$ and $\mathcal{F}^*$ can be proactively set up in Algorithm 1, it always can be ensured that $d^* \neq 0$. Owing to the relationship shown in (24), the normalized tensor variables $\bar{\mathcal{T}}(t)$, $\bar{\mathcal{Q}}(t) \in \mathbb{R}^{3\times3\times3}$ can be calculated by

$$\begin{aligned} \bar{\mathcal{T}}_{(j)} &= \frac{\mathcal{T}_{\lambda(j)}}{\sqrt{\frac{1}{2} \sum_{j=1}^{3} \sum_{k=1}^{3} \sum_{l=1}^{3} \mathcal{Q}_{\lambda(jkl)}^2}} \\ \bar{\mathcal{Q}}_{(j)} &= \frac{\mathcal{Q}_{\lambda(j)}}{\sqrt{\frac{1}{2} \sum_{j=1}^{3} \sum_{k=1}^{3} \sum_{l=1}^{3} \mathcal{Q}_{\lambda(jkl)}^2}}. \end{aligned} \tag{25}$$

Using (11), (19), (24), and (25), $\bar{\mathcal{T}}(t)$ and $\bar{\mathcal{Q}}(t)$ can be rewritten as $\bar{\mathcal{T}}_{(j)}(t) = \frac{\mathcal{T}_{(j)}(t)}{d^*}$ and $\bar{\mathcal{Q}}_{(j)}(t) = \frac{\mathcal{Q}_{(j)}(t)}{d^*}$. Moreover, according to (1) and (17), the normalized tensor variables $\bar{\mathcal{T}}(t)$ and $\bar{\mathcal{Q}}(t)$ can be formulated in terms of pose information and distance scale factor as follows:

$$\begin{aligned} \bar{\mathcal{T}}_{(j)} &= \frac{1}{d^*} \left({}^{c}R_{*(j)} {}^{d}t_{*}^{T} - {}^{c}t_{*} {}^{d}R_{*(j)}^{T} \right) \\ \bar{\mathcal{Q}}_{(j)} &= \frac{1}{d^*} {}^{c}R_{*(j)} {}^{d}t_{*}^{T} \left[{}^{d}R_{*(j)} \right]_{\times}, \end{aligned} \tag{26}$$

which indicates that

$$\begin{aligned} \bar{\mathcal{T}}_{(jl)} &= \frac{1}{d^*} \left({}^{d}t_{*(l)} {}^{c}R_{*(j)} - {}^{d}R_{*(lj)} {}^{c}t_{*} \right) \\ \bar{\mathcal{Q}}_{(jl)} &= \frac{1}{d^*} {}^{c}R_{*(j)} {}^{d}t_{*}^{T} \left[{}^{d}R_{*(j)} \right]_{\times(l)}. \end{aligned} \tag{27}$$

After taking the time derivative of (27) and using (5), the following expression can be obtained:

$$\begin{aligned} \dot{\bar{\mathcal{T}}}_{(jl)} &= \frac{1}{d^*} \left({}^{d}t_{*(l)} {}^{c}\dot{R}_{*(j)} - {}^{d}R_{*(lj)} {}^{c}\dot{t}_{*} \right) \\ &= \frac{{}^{d}R_{*(lj)}}{d^*} v + \left[\bar{\mathcal{T}}_{(jl)} \right]_{\times} \omega \\ \dot{\bar{\mathcal{Q}}}_{(jl)} &= \frac{1}{d^*} {}^{c}\dot{R}_{*(j)} {}^{d}t_{*}^{T} \left[{}^{d}R_{*(j)} \right]_{\times(l)} \\ &= \left[\bar{\mathcal{Q}}_{(jl)} \right]_{\times} \omega. \end{aligned} \tag{28}$$

This paper aims at utilizing the elements selected from the tensor variables to define the visual feedback. To achieve this objective, it is necessary to introduce the desired normalized tensor variables $\bar{\mathcal{T}}_d$, $\bar{\mathcal{Q}}_d \in \mathbb{R}^{3\times3\times3}$ corresponding to $\bar{\mathcal{T}}(t)$ and $\bar{\mathcal{Q}}(t)$, respectively. It is obvious that these two desired normalized tensor variables are defined in terms of the desired pose, reference pose, and current pose being equal to the desired pose. Hence, $\bar{\mathcal{T}}_d$ and $\bar{\mathcal{Q}}_d$ can be computed from the desired and reference images before the control task starts. Besides, according to the definition of desired normalized tensor variables, $\bar{\mathcal{T}}_d$ and $\bar{\mathcal{Q}}_d$ can be expressed with the desired pose information and distance scale factor, as follows:

$$\begin{aligned} \bar{\mathcal{T}}_{d(jl)} &= \frac{1}{d^*} \left({}^{d}t_{*(l)} {}^{d}R_{*(j)} - {}^{d}R_{*(lj)} {}^{d}t_{*} \right) \\ \bar{\mathcal{Q}}_{d(jl)} &= \frac{1}{d^*} {}^{d}R_{*(j)} {}^{d}t_{*}^{T} \left[{}^{d}R_{*(j)} \right]_{\times(l)}. \end{aligned} \tag{29}$$

4.3 System errors construction

From (28), it is clear that the time derivative of $\bar{\mathcal{T}}(t)$ is associated with the linear velocity $v(t)$ while the variation of $\bar{\mathcal{Q}}(t)$ is only related to the angular velocity $\omega(t)$. Therefore, a good choice is to define the translation and rotation errors with $\bar{\mathcal{T}}(t)$ and $\bar{\mathcal{Q}}(t)$, respectively. By considering the intrinsic relationship between the pose information and the tensor variables shown in (27) and (29), 3 elements of $\bar{\mathcal{T}}(t)$ and 6 elements of $\bar{\mathcal{Q}}(t)$ are chosen to construct the system errors $e_{\mathcal{T}}(t) \in \mathbb{R}^3$ and $e_{\mathcal{Q}}(t) \in \mathbb{R}^6$, respectively. Ideally, it would be preferred to design the rotation errors with only 3 elements of $\bar{\mathcal{Q}}(t)$, but according to the definition of $\bar{\mathcal{Q}}(t)$, any 3 elements of $\bar{\mathcal{Q}}(t)$ cannot be isomorphic to the rotation errors between the current and desired camera frames. More precisely, it can be obtained from (27) and (29) that

$$\bar{\mathcal{Q}}_{(jl)} - \bar{\mathcal{Q}}_{d(jl)} = \frac{1}{d^*} {}^{d}t_{*}^{T} \left[{}^{d}R_{*(j)} \right]_{\times(l)} \left({}^{c}R_{*(j)} - {}^{d}R_{*(j)} \right).$$

If ${}^{d}t_{*}^{T} \left[{}^{d}R_{*(j)} \right]_{\times(l)} \neq 0$, then $\bar{\mathcal{Q}}_{(jl)}(t) = \bar{\mathcal{Q}}_{d(jl)}$ indicates ${}^{c}R_{*(j)}(t) = {}^{d}R_{*(j)}$. Note that provided only 3 elements of ${}^{c}R_{*}(t)$ and ${}^{d}R_{*}$ are same, it cannot be concluded that these two rotation matrices are identical. Due to this fact, any 3 elements of $\bar{\mathcal{Q}}(t)$ are not isomorphic to the rotation errors, and 6 elements of $\bar{\mathcal{Q}}(t)$ are utilized in this paper to describe $e_{\mathcal{Q}}(t)$.

The selection criterion of the system errors is detailed in Algorithm 2. Specifically, if ${}^{d}R_{*(l1)} \neq 0$, then $\bar{\mathcal{T}}_{(1l)}(t)$ can be used to define the translation errors $e_{\mathcal{T}}(t)$. For the rotation errors, if $\bar{\mathcal{Q}}_{(1l_1)}^{T}(t) \bar{\mathcal{Q}}_{(1l_1)}(t) \neq \bar{\mathcal{Q}}_{(2l_2)}^{T}(t) \bar{\mathcal{Q}}_{(2l_2)}(t)$ and $\bar{\mathcal{Q}}_{(1l_1)}^{T}(t) \bar{\mathcal{Q}}_{(1l_1)}(t)$, $\bar{\mathcal{Q}}_{(2l_2)}^{T}(t) \bar{\mathcal{Q}}_{(2l_2)}(t) \neq 0$, then $e_{\mathcal{Q}}(t)$

Algorithm 2 System errors construction

Input:
Tensor variables, $\bar{\mathcal{T}}$, $\bar{\mathcal{Q}}$, $\bar{\mathcal{T}}_d$, $\bar{\mathcal{Q}}_d$;
Output:
System errors, $e_{\mathcal{T}}$, $e_{\mathcal{Q}}$;

for each $l \in \{1,2,3\}$ **do**
  if ${}^{d}R_{*(l1)} \neq 0$ **then**
    $e_{\mathcal{T}} \leftarrow \bar{\mathcal{T}}_{(1l)} - \bar{\mathcal{T}}_{d(1l)}$; Break;
  end if
end for
for each $l_1 \in \{1,2,3\}$ **do**
  for each $l_2 \in \{1,2,3\}$ **do**
    if $\bar{\mathcal{Q}}^T_{(1l_1)}\bar{\mathcal{Q}}_{(1l_1)} \neq \bar{\mathcal{Q}}^T_{(2l_2)}\bar{\mathcal{Q}}_{(2l_2)}$ and $\bar{\mathcal{Q}}^T_{(1l_1)}\bar{\mathcal{Q}}_{(1l_1)}, \bar{\mathcal{Q}}^T_{(2l_2)}\bar{\mathcal{Q}}_{(2l_2)} \neq 0$ **then**
      $e_{\mathcal{Q}} \leftarrow \begin{bmatrix}\bar{\mathcal{Q}}_{(1l_1)} - \bar{\mathcal{Q}}_{d(1l_1)} \\ \bar{\mathcal{Q}}_{(2l_2)} - \bar{\mathcal{Q}}_{d(2l_2)}\end{bmatrix}$; Break;
    end if
  end for
end for

can be constructed with $\bar{\mathcal{Q}}_{(1l_1)}(t)$ and $\bar{\mathcal{Q}}_{(2l_2)}(t)$. Based on (27), $\bar{\mathcal{Q}}_{1l_1}(t)$ and $\bar{\mathcal{Q}}_{1l_2}(t)$ can be related to the pose signals as follows:

$$\begin{aligned}\bar{\mathcal{Q}}_{(1l_1)} &= \frac{1}{d^*}\left({}^{d}t_*^T\left[{}^{d}R_{*(1)}\right]_{\times(l_1)}\right){}^{c}R_{*(1)} \\ \bar{\mathcal{Q}}_{(2l_2)} &= \frac{1}{d^*}\left({}^{d}t_*^T\left[{}^{d}R_{*(2)}\right]_{\times(l_2)}\right){}^{c}R_{*(2)}\end{aligned} \tag{30}$$

implying that

$$\begin{aligned}\bar{\mathcal{Q}}^T_{(1l_1)}\bar{\mathcal{Q}}_{(1l_1)} &= \left(\frac{1}{d^*}{}^{d}t_*^T\left[{}^{d}R_{*(1)}\right]_{\times(l_1)}\right)^2 \\ \bar{\mathcal{Q}}^T_{(2l_2)}\bar{\mathcal{Q}}_{(2l_2)} &= \left(\frac{1}{d^*}{}^{d}t_*^T\left[{}^{d}R_{*(2)}\right]_{\times(l_2)}\right)^2.\end{aligned}$$

Thus, $\bar{\mathcal{Q}}^T_{(1l_1)}(t)\bar{\mathcal{Q}}_{(1l_1)}(t) \neq \bar{\mathcal{Q}}^T_{(2l_2)}(t)\bar{\mathcal{Q}}_{(2l_2)}(t)$ and $\bar{\mathcal{Q}}^T_{(1l_1)}(t)\bar{\mathcal{Q}}_{(1l_1)}(t)$, $\bar{\mathcal{Q}}^T_{(2l_2)}(t)\bar{\mathcal{Q}}_{(2l_2)}(t) \neq 0$ indicate that ${}^{d}t_*^T[{}^{d}R_{*(1)}]_{\times(l_1)} \neq {}^{d}t_*^T[{}^{d}R_{*(2)}]_{\times(l_2)}$ and ${}^{d}t_*^T[{}^{d}R_{*(1)}]_{\times(l_1)}$, ${}^{d}t_*^T[{}^{d}R_{*(2)}]_{\times(l_2)} \neq 0$. As discussed in Section 3, the pose signals ${}^{d}R_*$ and ${}^{d}\bar{t}_*$ are selected proactively. Therefore, there always exist variables that satisfy $\bar{\mathcal{Q}}^T_{(1l_1)}(t)\bar{\mathcal{Q}}_{(1l_1)}(t) \neq \bar{\mathcal{Q}}^T_{(2l_2)}(t)\bar{\mathcal{Q}}_{(2l_2)}(t)$ and $\bar{\mathcal{Q}}^T_{(1l_1)}(t)\bar{\mathcal{Q}}_{(1l_1)}(t)$, $\bar{\mathcal{Q}}^T_{(2l_2)}(t)\bar{\mathcal{Q}}_{(2l_2)}(t) \neq 0$ by suitably choosing ${}^{d}R_*$ and ${}^{d}\bar{t}_*$.

To clearly explain why Algorithm 2 is applicable for the error definition, let us assume, for instance, that ${}^{d}R_{*(11)} \neq 0$, $\bar{\mathcal{Q}}^T_{(11)}(t)\bar{\mathcal{Q}}_{(11)}(t) \neq \bar{\mathcal{Q}}^T_{(21)}(t)\bar{\mathcal{Q}}_{(21)}(t)$, and $\bar{\mathcal{Q}}^T_{(11)}(t)\bar{\mathcal{Q}}_{(11)}(t)$, $\bar{\mathcal{Q}}^T_{(21)}(t)\bar{\mathcal{Q}}_{(21)}(t) \neq 0$ (i.e., ${}^{d}t_*^T\left[{}^{d}R_{*(1)}\right]_{\times(1)}$, ${}^{d}t_*^T\left[{}^{d}R_{*(2)}\right]_{\times(1)} \neq 0$). Then, $\bar{\mathcal{T}}_{(11)}(t)$, $\bar{\mathcal{Q}}_{(11)}(t)$, and $\bar{\mathcal{Q}}_{(21)}(t)$ are utilized to construct the system errors, i.e.,

$$\begin{aligned}e_{\mathcal{T}} &\triangleq \bar{\mathcal{T}}_{(11)} - \bar{\mathcal{T}}_{d(11)} = \begin{bmatrix}\bar{\mathcal{T}}_{(111)} - \bar{\mathcal{T}}_{d(111)} \\ \bar{\mathcal{T}}_{(121)} - \bar{\mathcal{T}}_{d(121)} \\ \bar{\mathcal{T}}_{(131)} - \bar{\mathcal{T}}_{d(131)}\end{bmatrix} \\ e_{\mathcal{Q}} &\triangleq \begin{bmatrix}\bar{\mathcal{Q}}_{(11)} - \bar{\mathcal{Q}}_{d(11)} \\ \bar{\mathcal{Q}}_{(21)} - \bar{\mathcal{Q}}_{d(21)}\end{bmatrix} = \begin{bmatrix}\bar{\mathcal{Q}}_{(111)} - \bar{\mathcal{Q}}_{d(111)} \\ \bar{\mathcal{Q}}_{(121)} - \bar{\mathcal{Q}}_{d(121)} \\ \bar{\mathcal{Q}}_{(131)} - \bar{\mathcal{Q}}_{d(131)} \\ \bar{\mathcal{Q}}_{(211)} - \bar{\mathcal{Q}}_{d(211)} \\ \bar{\mathcal{Q}}_{(221)} - \bar{\mathcal{Q}}_{d(221)} \\ \bar{\mathcal{Q}}_{(231)} - \bar{\mathcal{Q}}_{d(231)}\end{bmatrix}.\end{aligned} \tag{31}$$

According to (26), (29), and (31), $e_{\mathcal{T}}(t)$ and $e_{\mathcal{Q}}(t)$ can be rewritten as

$$\begin{aligned}e_{\mathcal{T}} &= \frac{1}{d^*}\left({}^{d}t_{*(1)}\left({}^{c}R_{*(1)} - {}^{d}R_{*(1)}\right) - {}^{d}R_{*(11)}\left({}^{c}t_* - {}^{d}t_*\right)\right) \\ e_{\mathcal{Q}} &= \frac{1}{d^*}\begin{bmatrix}{}^{d}t_*^T\left[{}^{d}R_{*(1)}\right]_{\times(1)}\left({}^{c}R_{*(1)} - {}^{d}R_{*(1)}\right) \\ {}^{d}t_i^T\left[{}^{d}R_{*(2)}\right]_{\times(1)}\left({}^{c}R_{*(2)} - {}^{d}R_{*(2)}\right)\end{bmatrix}.\end{aligned} \tag{32}$$

From (32), it can be concluded that if $e_{\mathcal{T}}(t)$, $e_{\mathcal{Q}}(t) \to 0$, then ${}^{c}t_*(t) \to {}^{d}t_*$ and $\left[{}^{c}R_{*(1)}(t) \;\; {}^{c}R_{*(2)}(t)\right] \to \left[{}^{d}R_{*(1)} \;\; {}^{d}R_{*(2)}\right]$ provided that ${}^{d}R_{*(11)} \neq 0$, ${}^{d}t_i^T\left[{}^{d}R_{*(1)}\right]_{\times(1)} \neq 0$, and ${}^{d}t_i^T\left[{}^{d}R_{*(2)}\right]_{\times(1)} \neq 0$. Since the third column of a rotation matrix can be represented by the cross-product of the other two, $\left[{}^{c}R_{*(1)}(t) \;\; {}^{c}R_{*(2)}(t)\right] \to \left[{}^{d}R_{*(1)} \;\; {}^{d}R_{*(2)}\right]$ indicates that ${}^{c}R_{*(3)}(t) \to {}^{d}R_{*(3)}$, i.e., ${}^{c}R_*(t) \to {}^{d}R_*$. Once ${}^{c}t_*(t) \to {}^{d}t_*$ and ${}^{c}R_*(t) \to {}^{d}R_*$, the control objective described in Section 2.1 is accomplished.

Note that although there might exist different tensor variables which fulfill the conditions presented in Algorithm 2, their latent meaning that facilitates the error definitions is actually the same. Let us take $\bar{\mathcal{Q}}_{(1l_1)}(t)$ as an example. It can be found from (30) that for different $\bar{\mathcal{Q}}_{(1l_1)}(t)$, the only distinction in their expression is the constant parameters ${}^{d}t_*^T\left[{}^{d}R_{*(1)}\right]_{\times(l_1)}$, while the time-varying parameter related to the rotation information, i.e., ${}^{c}R_{*(1)}(t)$, is identical. Actually, ${}^{c}R_{*(1)}(t)$ is the key that $\bar{\mathcal{Q}}_{(1l_1)}(t)$ can be used to quantify the rotation errors. Since all $\bar{\mathcal{Q}}_{(1l_1)}(t)$ contain the same effective information, choosing different $\bar{\mathcal{Q}}_{(1l_1)}(t)$ for error construction will not lead to difference in the control development. This also applies to $\bar{\mathcal{T}}_{(1l)}(t)$ and $\bar{\mathcal{Q}}_{(2l_2)}(t)$. Thus, the results obtained from (31) in the following are same to the ones based on other error definitions.

5 Control development

5.1 Adaptive control scheme

To design the controller, the derivation of system errors needs to be developed. From (28) and (31), the open-loop error system can be deduced, as follows:

$$\dot{e}_{\mathcal{T}} = \frac{{}^{d}R_{*(11)}}{d^*}v + \left[\bar{\mathcal{T}}_{(11)}\right]_{\times}\omega, \qquad \dot{e}_{\mathcal{Q}} = L_{\mathcal{Q}}\omega \tag{33}$$

with $L_{\mathcal{Q}}(t) \in \mathbb{R}^{6\times 3}$ given by

$$L_{\mathcal{Q}} \triangleq \begin{bmatrix} [\bar{\mathcal{Q}}_{(11)}]_\times \\ [\bar{\mathcal{Q}}_{(21)}]_\times \end{bmatrix} = \begin{bmatrix} 0 & -\bar{\mathcal{Q}}_{(131)} & \bar{\mathcal{Q}}_{(121)} \\ \bar{\mathcal{Q}}_{(131)} & 0 & -\bar{\mathcal{Q}}_{(111)} \\ -\bar{\mathcal{Q}}_{(121)} & \bar{\mathcal{Q}}_{(111)} & 0 \\ 0 & -\bar{\mathcal{Q}}_{(231)} & \bar{\mathcal{Q}}_{(221)} \\ \bar{\mathcal{Q}}_{(231)} & 0 & -\bar{\mathcal{Q}}_{(211)} \\ -\bar{\mathcal{Q}}_{(221)} & \bar{\mathcal{Q}}_{(211)} & 0 \end{bmatrix}. \tag{34}$$

Motivated by the open-loop error system presented in (33), the velocity inputs $v(t)$ and $\omega(t)$ are designed as follows:

$$\begin{aligned} v &= \frac{1}{{}^{d}R_{*(11)}} \left(-k_1 e_{\mathcal{T}} - \hat{d}^* \left[\bar{\mathcal{T}}_{(11)} \right]_\times \omega \right) \\ \omega &= -k_2 L_{\mathcal{Q}}^{+} e_{\mathcal{Q}} \end{aligned} \tag{35}$$

where k_1, $k_2 \in \mathbb{R}$ are positive constant gains, $L_{\mathcal{Q}}^{+}(t) \triangleq \left(L_{\mathcal{Q}}^T(t) L_{\mathcal{Q}}(t)\right)^{-1} L_{\mathcal{Q}}^T(t) \in \mathbb{R}^{3\times 6}$ is the pseudo-inverse of $L_{\mathcal{Q}}(t)$, and $\hat{d}^*(t) \in \mathbb{R}$ is the estimate of the unknown distance scale factor d^*. A property will be given in the following to show that $\left(L_{\mathcal{Q}}^T(t) L_{\mathcal{Q}}(t)\right)^{-1}$ is symmetric and positive definite, and thus the calculation of $L_{\mathcal{Q}}^{+}(t)$ is always feasible. Moreover, to compensate for the unknown distance information, the update law for $\hat{d}^*(t)$ is given by

$$\dot{\hat{d}}^* = k_3 e_{\mathcal{T}}^T \left[\bar{\mathcal{T}}_{(11)} \right]_\times \omega \tag{36}$$

with $k_3 \in \mathbb{R}$ being a positive constant gain.

After substituting (35) into (33), the closed-loop error system can be obtained, as follows:

$$\begin{aligned} \dot{e}_{\mathcal{T}} &= -\frac{1}{d^*} k_1 e_{\mathcal{T}} + \frac{\tilde{d}^*}{d^*} \left[\bar{\mathcal{T}}_{(11)} \right]_\times \omega \\ \dot{e}_{\mathcal{Q}} &= -k_2 L_{\mathcal{Q}} L_{\mathcal{Q}}^{+} e_{\mathcal{Q}} \end{aligned} \tag{37}$$

where $\tilde{d}^*(t) \triangleq d^* - \hat{d}^*(t) \in \mathbb{R}$ is the estimation error of d^*. The control inputs in (35) and the adaptation mechanism in (36) are developed via an iterative analysis procedure. The main idea is to appropriately design $v(t)$, $\omega(t)$ along with the update law for $\hat{d}^*(t)$ to shape the closed-loop error system into the desired form presented in (37). This desired form is deduced through Lyapunov-based methods and will be used to facilitate the following stability analysis. Similar techniques can be found in Dixon et al. (2001); Chen et al. (2005). Moreover, the adaptation mechanism in (36) is utilized to actively compensate for the unknown distance information and to make the system errors converge to zero, instead of guaranteeing $\hat{d}^*(t)$ converges to d^*. Actually, conventional adaptive control methods do not ensure or require convergence of the parameter estimates to their corresponding true values (Slotine and Li, 1991).

5.2 Stability analysis

In this section, a property about the interaction matrix $L_{\mathcal{Q}}(t)$ is presented firstly, and then the stability is conducted based on this property and Lyapunov based techniques.

Property 1. *The matrix $\left(L_{\mathcal{Q}}^T(t) L_{\mathcal{Q}}(t)\right)^{-1}$ is symmetric and positive definite.*

Proof. See Appendix B. □

Theorem 1. *Consider the system (33) under the adaptive control inputs designed in (35) and the update law for the unknown distance designed in (36). Then,*

1) *The equilibria of the closed-loop system (37) are given by*

$$\begin{aligned} &\Omega_1 : \begin{bmatrix} e_{\mathcal{T}} \\ e_{\mathcal{Q}} \end{bmatrix} = 0_9, \qquad \Omega_2 : \begin{bmatrix} e_{\mathcal{T}} \\ e_{\mathcal{Q}} \end{bmatrix} = -2 \begin{bmatrix} 0_3 \\ \bar{\mathcal{Q}}_{d(11)} \\ \bar{\mathcal{Q}}_{d(21)} \end{bmatrix}, \\ &\Omega_3 : \begin{bmatrix} e_{\mathcal{T}} \\ e_{\mathcal{Q}} \end{bmatrix} = -2 \begin{bmatrix} 0_3 \\ \bar{\mathcal{Q}}_{d(11)} \\ 0_3 \end{bmatrix}, \ \Omega_4 : \begin{bmatrix} e_{\mathcal{T}} \\ e_{\mathcal{Q}} \end{bmatrix} = -2 \begin{bmatrix} 0_3 \\ 0_3 \\ \bar{\mathcal{Q}}_{d(21)} \end{bmatrix}. \end{aligned} \tag{38}$$

2) *The equilibria Ω_2, Ω_3, and Ω_4 are unstable, and the desired configuration Ω_1 is almost globally asymptotically stable.*

Proof. A non-negative function $V(t) \in \mathbb{R}$ is introduced to prove Theorem 1, which is defined as

$$V \triangleq \frac{1}{2} d^* e_{\mathcal{T}}^T e_{\mathcal{T}} + \frac{1}{2} e_{\mathcal{Q}}^T e_{\mathcal{Q}} + \frac{1}{2k_3} \tilde{d}^{*2}. \tag{39}$$

After taking the time derivative of (39) and exploiting (36) and (37), it can be obtained that

$$\begin{aligned} \dot{V} &= e_{\mathcal{T}}^T d^* \dot{e}_{\mathcal{T}} + e_{\mathcal{Q}}^T \dot{e}_{\mathcal{Q}} - \tilde{d}^* \frac{\dot{\hat{d}}^*}{k_3} \\ &= -k_1 e_{\mathcal{T}}^T e_{\mathcal{T}} - k_2 e_{\mathcal{Q}}^T L_{\mathcal{Q}} L_{\mathcal{Q}}^{+} e_{\mathcal{Q}}. \end{aligned} \tag{40}$$

Furthermore, using $L_{\mathcal{Q}}^{+}(t) = \left(L_{\mathcal{Q}}^T(t) L_{\mathcal{Q}}(t)\right)^{-1} L_{\mathcal{Q}}^T(t)$ and Property 1, it can be concluded that

$$\dot{V} = -k_1 e_{\mathcal{T}}^T e_{\mathcal{T}} - k_2 e_{\mathcal{Q}}^{'T} \left(L_{\mathcal{Q}}^T L_{\mathcal{Q}}\right)^{-1} e_{\mathcal{Q}}^{'} \leq 0 \tag{41}$$

where $e_{\mathcal{Q}}^{'}(t) \triangleq L_{\mathcal{Q}}^T(t) e_{\mathcal{Q}}(t) \in \mathbb{R}^3$. Based on (39) and (41), it can be concluded that $e_{\mathcal{T}}(t)$, $e_{\mathcal{Q}}(t)$, $\tilde{d}^* \in \mathcal{L}_\infty$ and $e_{\mathcal{T}}(t)$, $e_{\mathcal{Q}}^{'}(t) \in \mathcal{L}_2$. Then, using (31) and the facts that $\bar{\mathcal{T}}_{d(11)}$, $\bar{\mathcal{Q}}_{d(11)}$, and $\bar{\mathcal{Q}}_{d(21)}$ are bounded, $\bar{\mathcal{T}}_{(11)}(t)$, $\bar{\mathcal{Q}}_{(11)}(t)$, $\bar{\mathcal{Q}}_{(21)}(t) \in \mathcal{L}_\infty$ can be obtained. With the aforementioned bounded variables, it can be determined from (33), (34), and (35) that $v(t)$, $\omega(t)$, $L_{\mathcal{Q}}(t)$, $\dot{e}_{\mathcal{T}}(t)$, $\dot{e}_{\mathcal{Q}}(t) \in \mathcal{L}_\infty$. Since $L_{\mathcal{Q}}(t)$, $\dot{e}_{\mathcal{Q}}(t) \in \mathcal{L}_\infty$, $\dot{e}_{\mathcal{Q}}^{'}(t) \in \mathcal{L}_\infty$ can be deduced. Moreover, as $e_{\mathcal{T}}(t)$, $e_{\mathcal{Q}}^{'}(t) \in \mathcal{L}_2$ and $\dot{e}_{\mathcal{T}}(t)$, $\dot{e}_{\mathcal{Q}}^{'}(t) \in \mathcal{L}_\infty$, Barbalat's lemma (Khalil, 2002) can be exploited to infer that $\lim_{t\to\infty} e_{\mathcal{T}}(t), e_{\mathcal{Q}}^{'}(t) = 0$.

It can be found from Property 1 that the rank of $L_{\mathcal{Q}}^T(t)$ is three, and there exists a null space $\ker\left(L_{\mathcal{Q}}^T(t)\right)$ such that $\forall u \in \ker\left(L_{\mathcal{Q}}^T(t)\right)$, $L_{\mathcal{Q}}^T u = 0$. Therefore, $\lim_{t\to\infty} e_{\mathcal{Q}}^{'}(t) = \lim_{t\to\infty} L_{\mathcal{Q}}^T(t) e_{\mathcal{Q}}(t) = 0$ does not implies that $\lim_{t\to\infty} e_{\mathcal{Q}}(t) = 0$, i.e., the closed-loop system (37) may contain multiple equilibria. It is clear that the desired configuration Ω_1 is one of the equilibria. In the following, it is proved that Ω_2, Ω_3, and Ω_4 are also equilibria with respect to (37).

Since $L_{\mathcal{Q}}^T(t) \in \mathbb{R}^{3\times 6}$ has rank three, the dimension of its null space is three. Based on (34) and the facts that $\left[\bar{\mathcal{Q}}_{(11)}(t)\right]_\times \bar{\mathcal{Q}}_{(11)}(t) = \left[\bar{\mathcal{Q}}_{(21)}(t)\right]_\times \bar{\mathcal{Q}}_{(21)}(t) = 0_3$, and $\left[\bar{\mathcal{Q}}_{(11)}(t)\right]_\times \bar{\mathcal{Q}}_{(21)}(t) = -\left[\bar{\mathcal{Q}}_{(21)}(t)\right]_\times \bar{\mathcal{Q}}_{(11)}(t)$, a basis of the null space $\ker\left(L_{\mathcal{Q}}^T(t)\right)$ can be obtained, as

follows:

$$u_1 = \begin{bmatrix} \bar{\mathcal{Q}}_{(11)} \\ 0_3 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0_3 \\ \bar{\mathcal{Q}}_{(21)} \end{bmatrix}, \quad u_3 = \begin{bmatrix} \bar{\mathcal{Q}}_{(21)} \\ \bar{\mathcal{Q}}_{(11)} \end{bmatrix}. \tag{42}$$

If $e_{\mathcal{Q}}(t) \in \ker\left(L_{\mathcal{Q}}^T(t)\right)$, then it can be concluded that

$$e_{\mathcal{Q}} = \gamma_1 u_1 + \gamma_2 u_2 + \gamma_3 u_3 \tag{43}$$

where γ_1, γ_2, $\gamma_3 \in \mathbb{R}$ are constants that cannot be zero simultaneously. According to (31) and (42), (43) can be rewritten as

$$\begin{bmatrix} \bar{\mathcal{Q}}_{d(11)} \\ \bar{\mathcal{Q}}_{d(21)} \end{bmatrix} = -\begin{bmatrix} (\gamma_1 - 1)\,\bar{\mathcal{Q}}_{(11)} + \gamma_3 \bar{\mathcal{Q}}_{(21)} \\ (\gamma_2 - 1)\,\bar{\mathcal{Q}}_{(21)} + \gamma_3 \bar{\mathcal{Q}}_{(11)} \end{bmatrix}. \tag{44}$$

Furthermore, by using (26), (29), and the facts that ${}^cR_{*(1)}^T(t)\,{}^cR_{*(2)}(t) = {}^dR_{*(1)}^T\,{}^dR_{*(2)} = 0$, ${}^cR_{*(j)}^T(t)\,{}^cR_{*(j)}(t) = {}^dR_{*(j)}^T\,{}^dR_{*(j)} = 1$, the following properties can be derived:

$$\bar{\mathcal{Q}}_{d(11)}^T \bar{\mathcal{Q}}_{d(21)} = 0,\ \bar{\mathcal{Q}}_{d(11)}^T \bar{\mathcal{Q}}_{d(11)} = a_1,\ \bar{\mathcal{Q}}_{d(21)}^T \bar{\mathcal{Q}}_{d(21)} = a_2 \tag{45}$$

and

$$\bar{\mathcal{Q}}_{(11)}^T \bar{\mathcal{Q}}_{(21)} = 0,\ \ \bar{\mathcal{Q}}_{(11)}^T \bar{\mathcal{Q}}_{(11)} = a_1,\ \ \bar{\mathcal{Q}}_{(21)}^T \bar{\mathcal{Q}}_{(21)} = a_2 \tag{46}$$

where a_1, $a_2 \in \mathbb{R}$ are positive constants defined as

$$\begin{aligned} a_1 &\triangleq \left(\frac{1}{d^*}\,{}^d t_*^T \left[{}^dR_{*(1)}\right]_{\times(1)}\right)^2 \\ a_2 &\triangleq \left(\frac{1}{d^*}\,{}^d t_*^T \left[{}^dR_{*(2)}\right]_{\times(1)}\right)^2. \end{aligned} \tag{47}$$

Note that the selection criterion described in Section 4.3 ensures that $\bar{\mathcal{Q}}_{(11)}^T(t)\bar{\mathcal{Q}}_{(11)}(t) \neq \bar{\mathcal{Q}}_{(21)}^T(t)\bar{\mathcal{Q}}_{(21)}(t)$, and hence $a_1 \neq a_2$. Substituting (44) into (45) and using (46) to collect similar terms, it can be determined that

$$\begin{aligned} &(\gamma_1 - 1)\,\gamma_3 a_1 + (\gamma_2 - 1)\,\gamma_3 a_2 = 0, \\ &(\gamma_1 - 1)^2\, a_1 + \gamma_3^2 a_2 = a_1, \\ &(\gamma_2 - 1)^2\, a_2 + \gamma_3^2 a_1 = a_2. \end{aligned} \tag{48}$$

According to the last two equations of (48), the terms related to γ_3 are eliminated, and it can be obtained that

$$\left(\gamma_1^2 - 2\gamma_1\right) a_1^2 - \left(\gamma_2^2 - 2\gamma_2\right) a_2^2 = 0. \tag{49}$$

If $\gamma_3 \neq 0$, then based on $a_1 \neq a_2$, it can be concluded that γ_1 and γ_2 satisfying the first equation of (48) and (49) do not exist. If $\gamma_3 = 0$, then it can be obtained from (48) that $(\gamma_1 - 1)^2\, a_1 = a_1$ and $(\gamma_2 - 1)^2\, a_2 = a_2$. Based on these two equations, it can be concluded that $(\gamma_1, \gamma_2, \gamma_3)$ can be selected as $(2, 2, 0)$, $(2, 0, 0)$, or $(0, 2, 0)$. Therefore, there exist another three equilibria for the closed-loop system (37). Specifically, based on (44), the corresponding values of $\bar{\mathcal{Q}}_{(11)}(t)$ and $\bar{\mathcal{Q}}_{(21)}(t)$ for the equilibria Ω_2, Ω_3, and Ω_4 are given respectively by

$$\begin{aligned} \begin{bmatrix} \bar{\mathcal{Q}}_{(11)} \\ \bar{\mathcal{Q}}_{(21)} \end{bmatrix} &= -\begin{bmatrix} \bar{\mathcal{Q}}_{d(11)} \\ \bar{\mathcal{Q}}_{d(21)} \end{bmatrix}, \\ \begin{bmatrix} \bar{\mathcal{Q}}_{(11)} \\ \bar{\mathcal{Q}}_{(21)} \end{bmatrix} &= \begin{bmatrix} -\bar{\mathcal{Q}}_{d(11)} \\ \bar{\mathcal{Q}}_{d(21)} \end{bmatrix}, \\ \begin{bmatrix} \bar{\mathcal{Q}}_{(11)} \\ \bar{\mathcal{Q}}_{(21)} \end{bmatrix} &= \begin{bmatrix} \bar{\mathcal{Q}}_{d(11)} \\ -\bar{\mathcal{Q}}_{d(21)} \end{bmatrix}. \end{aligned} \tag{50}$$

According to $\lim_{t\to\infty} e_{\mathcal{T}}(t) = 0$, (31), and (50), the equilibria Ω_2, Ω_3, and Ω_4 given in (38) can be obtained.

Now, the proof of the second claim of Theorem 1 is presented. Firstly, we show that Ω_2 is unstable. Let $V_{\Omega_2}(t) \in \mathbb{R}$ be a non-negative function defined as

$$V_{\Omega_2} \triangleq \frac{1}{2}\left(e_{\mathcal{Q}} + 2\begin{bmatrix} \bar{\mathcal{Q}}_{d(11)} \\ \bar{\mathcal{Q}}_{d(21)} \end{bmatrix}\right)^T \left(e_{\mathcal{Q}} + 2\begin{bmatrix} \bar{\mathcal{Q}}_{d(11)} \\ \bar{\mathcal{Q}}_{d(21)} \end{bmatrix}\right). \tag{51}$$

$V_{\Omega_2}(t) = 0$ if and only if $e_{\mathcal{Q}}(t)$ reaches the equilibrium Ω_2. Taking the time derivative of (51) and using (31), (37), and Property 1, it can be determined that

$$\dot{V}_{\Omega_2} = k_2\left(L_{\mathcal{Q}}^T \begin{bmatrix} \bar{\mathcal{Q}}_{d(11)} \\ \bar{\mathcal{Q}}_{d(21)} \end{bmatrix}\right)^T \left(L_{\mathcal{Q}}^T L_{\mathcal{Q}}\right)^{-1} \left(L_{\mathcal{Q}}^T \begin{bmatrix} \bar{\mathcal{Q}}_{d(11)} \\ \bar{\mathcal{Q}}_{d(21)} \end{bmatrix}\right) \geq 0. \tag{52}$$

Based on (51) and (52), Lyapunov's first instability theorem (Khalil, 2002) can be used to conclude that the origin such that $V_{\Omega_2}(t) = 0$ is unstable, i.e., Ω_2 is unstable.

Secondly, let us show that Ω_3 and Ω_4 are unstable. Since the non-negative functions, which are similar to (51), cannot be found for Ω_3 and Ω_4, Lyapunov's first instability theorem is no longer applicable. To facilitate the following analysis, Chetaev's theorem (Khalil, 2002) is utilized. The Chetaev function $V_{\Omega_3}(t) \in \mathbb{R}$ with respect to Ω_3 is defined as

$$\begin{aligned} V_{\Omega_3} \triangleq\ & \frac{1}{2}\left(\bar{\mathcal{Q}}_{(11)} + \bar{\mathcal{Q}}_{d(11)}\right)^T \left(\bar{\mathcal{Q}}_{(11)} + \bar{\mathcal{Q}}_{d(11)}\right) \\ & - \frac{1}{2}\left(\bar{\mathcal{Q}}_{(21)} - \bar{\mathcal{Q}}_{d(21)}\right)^T \left(\bar{\mathcal{Q}}_{(21)} - \bar{\mathcal{Q}}_{d(21)}\right). \end{aligned} \tag{53}$$

Define the set $B_r \triangleq \left\{x \in \mathbb{R}^6 \mid \|x\| < r\right\}$ with $x(t) \triangleq \left[\left(\bar{\mathcal{Q}}_{(11)}(t) + \bar{\mathcal{Q}}_{d(11)}\right)^T \ \ \left(\bar{\mathcal{Q}}_{(21)}(t) - \bar{\mathcal{Q}}_{d(21)}\right)^T\right]^T$ and r being an arbitrarily small positive constant. If $\bar{\mathcal{Q}}_{(11)}(t) = -\bar{\mathcal{Q}}_{d(11)}$ and $\bar{\mathcal{Q}}_{(21)}(t) = \bar{\mathcal{Q}}_{d(21)}$, i.e., $e_{\mathcal{Q}}(t)$ is equal to the value at Ω_3, then $x(t) = 0_6$ and $V_{\Omega_3}(t) = 0$. Let us also define a subset of B_r such that $U \triangleq \left\{x \in \mathbb{R}^6 \mid V_{\Omega_3} > 0\right\}$. Based on (53), it can be found that if $\bar{\mathcal{Q}}_{(21)}(t) - \bar{\mathcal{Q}}_{d(21)} = 0_3$, $V_{\Omega_3}(t) > 0$ at points arbitrarily close to the origin $x(t) = 0_6$. Provided that $V_{\Omega_3}(t) = 0$ at $x(t) = 0_6$ and $V_{\Omega_3}(t) > 0$ for some $x(t)$ arbitrarily close to the origin, the set U can be always constructed (Khalil, 2002). By utilizing (31), (37), and Property 1, the time derivative of $V_{\Omega_3}(t)$ can be derived as follows:

$$\dot{V}_{\Omega_3} = k_2\left(L_{\mathcal{Q}}^T \begin{bmatrix} \bar{\mathcal{Q}}_{d(11)} \\ \bar{\mathcal{Q}}_{d(21)} \end{bmatrix}\right)^T \left(L_{\mathcal{Q}}^T L_{\mathcal{Q}}\right)^{-1} \left(L_{\mathcal{Q}}^T \begin{bmatrix} \bar{\mathcal{Q}}_{d(11)} \\ \bar{\mathcal{Q}}_{d(21)} \end{bmatrix}\right) \geq 0. \tag{54}$$

It can be concluded from (54) that $V_{\Omega_3}(t) \geq V_{\Omega_3}(0)$. Since $V_{\Omega_3}(t)$ is bounded on U and $\dot{V}_{\Omega_3}(t) > 0$ everywhere in U, $x(t)$ must leave U after picking the initial condition $x(0) \in U$. Moreover, since $V_{\Omega_3}(t) \geq V_{\Omega_3}(0)$, $x(t)$ must leave U through the sphere $\|x(t)\| = r$ and not through the edges $V_{\Omega_3}(t) = 0$, e.g., Ω_3. According to the above arguments, Chetaev's theorem can be exploited to obtain that Ω_3 is unstable. Finally, consider the following Chetaev function $V_{\Omega_4}(t) \in \mathbb{R}$:

$$\begin{aligned} V_{\Omega_4} \triangleq\ & -\frac{1}{2}\left(\bar{\mathcal{Q}}_{(11)} - \bar{\mathcal{Q}}_{d(11)}\right)^T \left(\bar{\mathcal{Q}}_{(11)} - \bar{\mathcal{Q}}_{d(11)}\right) \\ & + \frac{1}{2}\left(\bar{\mathcal{Q}}_{(21)} + \bar{\mathcal{Q}}_{d(21)}\right)^T \left(\bar{\mathcal{Q}}_{(21)} + \bar{\mathcal{Q}}_{d(21)}\right) \end{aligned} \tag{55}$$

whose time derivative is given by

$$\dot{V}_{\Omega_4} = k_2 \left(L_{\mathcal{Q}}^T \begin{bmatrix} \bar{\mathcal{Q}}_{d(11)} \\ \bar{\mathcal{Q}}_{d(21)} \end{bmatrix} \right)^T \left(L_{\mathcal{Q}}^T L_{\mathcal{Q}} \right)^{-1} \left(L_{\mathcal{Q}}^T \begin{bmatrix} \bar{\mathcal{Q}}_{d(11)} \\ \bar{\mathcal{Q}}_{d(21)} \end{bmatrix} \right) \geq 0. \tag{56}$$

Based on (55) and (56), similar analysis techniques can be utilized to conclude that Ω_4 is also unstable.

Finally, since Ω_2, Ω_3, and Ω_4 are unstable, it can be obtained that Ω_1 is almost globally attractive (Chaturvedi et al., 2011; Tayebi et al., 2013). □

More discussion about these four equilibria is given now.

If $e_{\mathcal{Q}}(0) \in \Phi = \left\{ e_{\mathcal{Q}} \in \mathbb{R}^6 | e_{\mathcal{Q}}^T e_{\mathcal{Q}} < 4 \min\{a_1, a_2\} \right\}$, then the system errors converge to Ω_1 asymptotically without restricting by other equilibria. To prove this claim, let us define $V_{\mathcal{Q}}(t) \triangleq e_{\mathcal{Q}}^T(t) e_{\mathcal{Q}}(t)$. According to (38) and (45), it can be obtained that the corresponding values of $V_{\mathcal{Q}}(t)$ at Ω_2, Ω_3, and Ω_4 are $4(a_1 + a_2)$, $4a_1$, and $4a_2$, respectively. Therefore, the equilibrium in the set Φ is nothing else but Ω_1. Furthermore, the time derivative of $V_{\mathcal{Q}}(t)$ along (37) is given by

$$\dot{V}_{\mathcal{Q}} = -2k_2 e_{\mathcal{Q}}^{'T} \left(L_{\mathcal{Q}}^T L_{\mathcal{Q}} \right)^{-1} e_{\mathcal{Q}}^{'} \leq 0 \tag{57}$$

where $e_{\mathcal{Q}}^{'}(t)$ is defined after (41). (57) implies that $\forall t \geq 0$, $V_{\mathcal{Q}}(t) \leq V_{\mathcal{Q}}(0)$. Based on $V_{\mathcal{Q}}(t) \leq V_{\mathcal{Q}}(0)$, it can be concluded that if $e_{\mathcal{Q}}(0) \in \Phi$, then $\forall t \geq 0$, $e_{\mathcal{Q}}(t) \in \Phi$, i.e., Φ is a positively invariant set. Since Ω_1 is the only equilibrium in the positively invariant set Φ, the system errors converge to Ω_1 provided $e_{\mathcal{Q}}(0) \in \Phi$.

Moreover, it is interesting to note that from (26), (29), and (50), the corresponding value of ${}^cR_*(t)$ for the equilibria Ω_2, Ω_3, and Ω_4 can be derived respectively as follows:

$$\begin{aligned} {}^cR_* &= \begin{bmatrix} -{}^dR_{*(1)} & -{}^dR_{*(2)} & {}^dR_{*(3)} \end{bmatrix}, \\ {}^cR_* &= \begin{bmatrix} -{}^dR_{*(1)} & {}^dR_{*(2)} & -{}^dR_{*(3)} \end{bmatrix}, \\ {}^cR_* &= \begin{bmatrix} {}^dR_{*(1)} & -{}^dR_{*(2)} & -{}^dR_{*(3)} \end{bmatrix}. \end{aligned} \tag{58}$$

From a geometric point of view, (58) indicates that the rotation of the current camera frame $\mathcal{F}_c$ at Ω_2, Ω_3, and Ω_4 differs from the desired frame $\mathcal{F}_d$ by 180° of rotation about the axes ${}^dR_{*(3)}$, ${}^dR_{*(2)}$, and ${}^dR_{*(1)}$, respectively. For instance, if ${}^dR_*$ is chosen as the identity matrix, then Ω_2, Ω_3, and Ω_4 correspond to $\mathcal{F}_c$ differing from $\mathcal{F}_d$ by 180° of rotation about the z-axis (i.e., optical axis), y-axis, and x-axis, respectively. Furthermore, in the case that ${}^dR_* = {}^dR_i$, if the rotation between the initial and desired camera frames, i.e., $\mathcal{F}_i$ and $\mathcal{F}_d$, is equal to 180° along the z-axis, or y-axis, or x-axis, then the system will be subject to the unstable equilibrium Ω_2, or Ω_3, or Ω_4 at the beginning. Note that under most circumstances, a 180° rotation along the y-axis or x-axis will make the target object out of the camera field of view, and thus Ω_3 and Ω_4 usually do not belong to the system workspace.

6 Simulation and experimental results

Simulation and experiment studies are performed with the aid of the open-source ViSP library (Marchand et al., 2005) to illustrate the performance of the proposed approach. All the experiments are conducted by making use of a grayscale camera attached to the end-effector of a 6 DOF robot. The camera provides images with a frame rate of 30 fps and a resolution of 640×480 px. Besides, the adjustable parameters ${}^dR_*$ and ${}^d\bar{t}_*$ in Algorithm 1 are designed as ${}^dR_* = {}^dR_i$ and ${}^d\bar{t}_* = -0.5\,{}^d\bar{t}_i$. $\bar{\mathcal{T}}_{(11)}(t)$, $\bar{\mathcal{Q}}_{(11)}(t)$, and $\bar{\mathcal{Q}}_{(21)}(t)$ are used to define the system errors.

6.1 Comparison validation

Comparison results are provided to evaluate the effectiveness of the proposed approach. Existing four methods are selected to facilitate the comparison, and brief introduction of these four methods is listed as follows:

- *Method 1 (Silveira and Malis, 2012):* Instead of using trifocal tensor elements, this method combines one feature point with homography elements to define visual features and control inputs.
- *Method 2:* All elements of the normalized trifocal tensor $\bar{\mathcal{T}}(t)$ are used to define the visual features, and then the control inputs are generated with the strategy described in Section 2.3. Note that the visual feedback is defined with $\bar{\mathcal{T}}(t)$, which ensures the interaction matrix $L(t)$ depends on the unknown constant distance d^*. Since d^* is invariable, an approximation of d^*, i.e., $\hat{d}^* = 0.4$, has been selected to compute the pseudo-inverse of $L(t)$ involved in (10).
- *Method 3:* SVD-based techniques are exploited to obtain explicit pose information (i.e., rotation and scaled translation) from the trifocal tensor, and then the controller is developed with the obtained pose signals. Since the decomposed translation is scaled by an unknown constant, an approximation is also chosen to facilitate the control input calculation.
- *Method 4 (Malis and Chaumette, 2000):* The homography-based 2.5 D method combines one feature point with rotation information to construct visual features and control inputs.

Table 1. Errors between the initial frame $\mathcal{F}_i$ and the desired frame $\mathcal{F}_d$ for a dot scene.

Case	Translation Errors (m)			Rotation Errors ($^\circ$)		
	t_x	t_y	t_z	r_x	r_y	r_z
Pure Translation	-0.30	0.11	0.39	0	0	0
Large Rotation	-0.08	0.04	0.37	0	0	-120
General Motion	-0.57	0.53	0.30	31	24	27

In this set of experiments, the regulation of nine feature points are considered. These tracked points are non-coplanar black dots with white background. Thanks to the high contrast between the black features and white surface, the detection and tracking of the nine dots can be obtained easily through the blob tracker provided by ViSP. In fact, this experimental setting is aimed at reproducing the identical initial condition to conduct the comparison experiment. Regulating a set of points belonging to less structured objects will be studied in Section 6.3 to further show the viability of developed approach in more realistic scene. In the following, three cases with different initial and desired camera poses are utilized to test the methods.

- *Pure Translation Case:* The first experiment considers a pure translation along the three axes. The errors between the initial frame $\mathcal{F}_i$ and the desired frame $\mathcal{F}_d$ are reported in

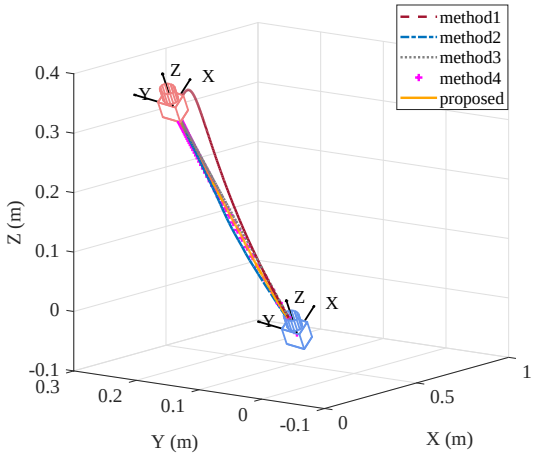

Figure 3. Cartesian space motion trajectories during the pure translation case. The blue and red cameras represent the initial and desired poses, respectively.

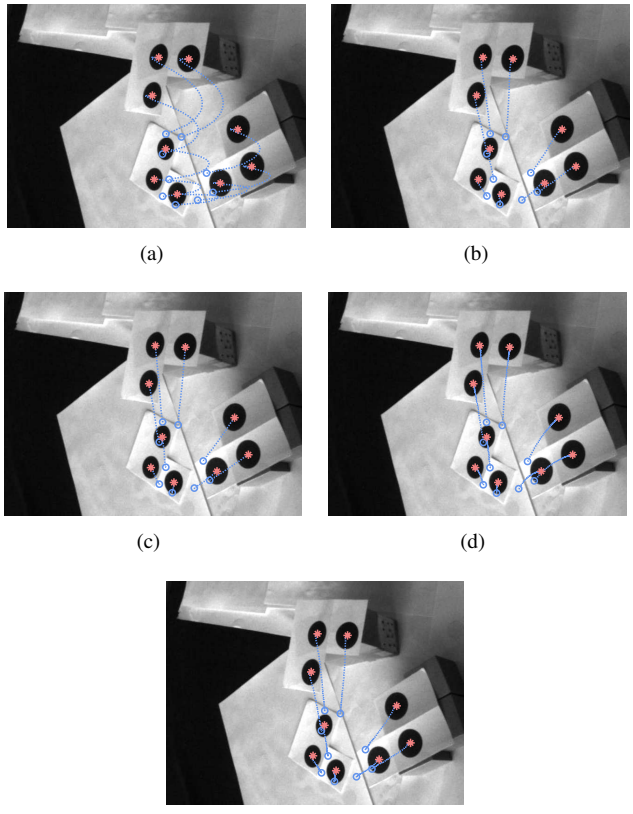

Figure 4. Experimental comparisons of the image trajectory for: (a) method 1, (b) method 2, (c) method 3, (d) method 4, (e) and proposed approach in the pure translation case. The circles and asterisks represent the initial and desired position of feature points, respectively.

Table 1. Meanwhile, the corresponding experimental results are shown in Figures 3, 4, and 5. From the results, it can be found that the motion and image trajectories of Method 2, Method 3, Method 4, and proposed approach are linear without redundant rotation component. In fact, for the pure translation case, the linear image trajectories can avoid losing the feature points effectively. However, as presented in Figure 5(a), Method 1 introduces superfluous

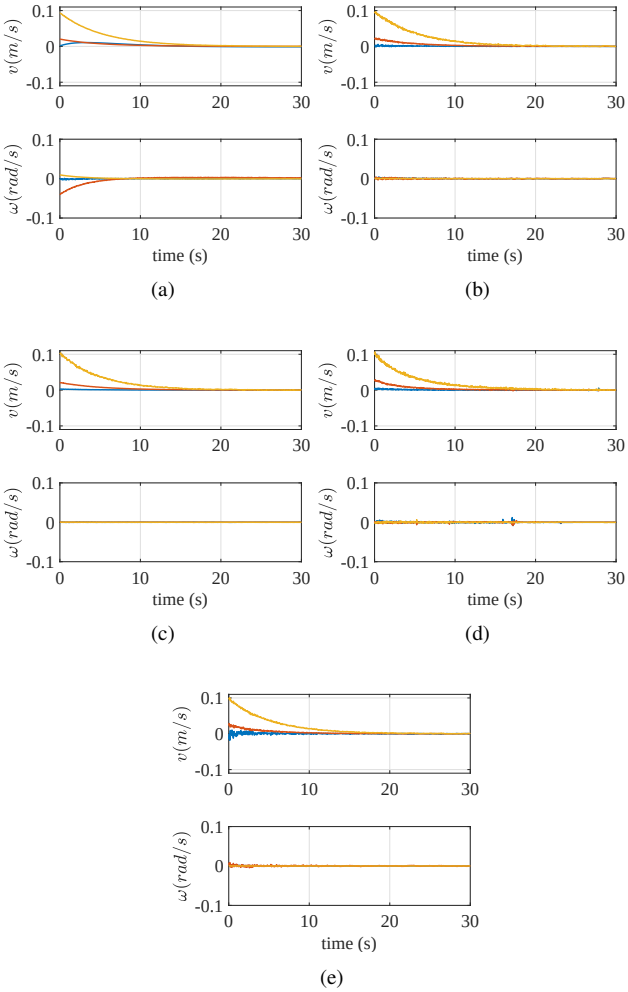

Figure 5. Experimental comparisons of the control inputs for: (a) method 1, (b) method 2, (c) method 3, (d) method 4, (e) and proposed approach in the pure translation case.

rotation movement, which results in curved motion and image trajectories. Figure 5(e) shows that the control inputs of the proposed approach contain larger variations than other methods at the beginning. The proposed controller is constructed directly with some of the tensor elements. Since the camera will jitter when starting up the robot, the estimation stability of these tensor elements will be affected and that leads to the input jitter at the beginning. Nevertheless, the regulation error can be reduced by the proposed approach during the motion procedure, and the convergence can be achieved in both Cartesian and image space.

• *Large Rotation Case:* For visual servoing, one of the most challenging configurations is the large rotation error around the z-axis. To further evaluate the proposed approach, a 120° z-axis rotation is considered here. The experimental results are depicted in Figures 6, 7, and 8. From Figure 7(e), it can be seen that the image trajectories of the feature points corresponding to the proposed approach follow a spiral motion, which is exactly as expected due to the rotation motion around the z-th axis. Although all image trajectories are curved, no feature point moves too near to the image boarders. Note that only the rotation error along the z-axis

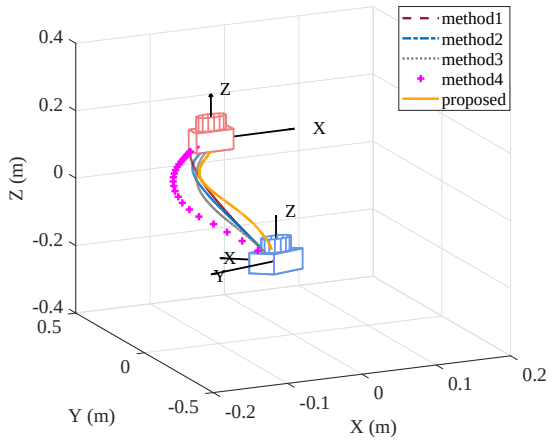

Figure 6. Cartesian space motion trajectories during the large rotation case.

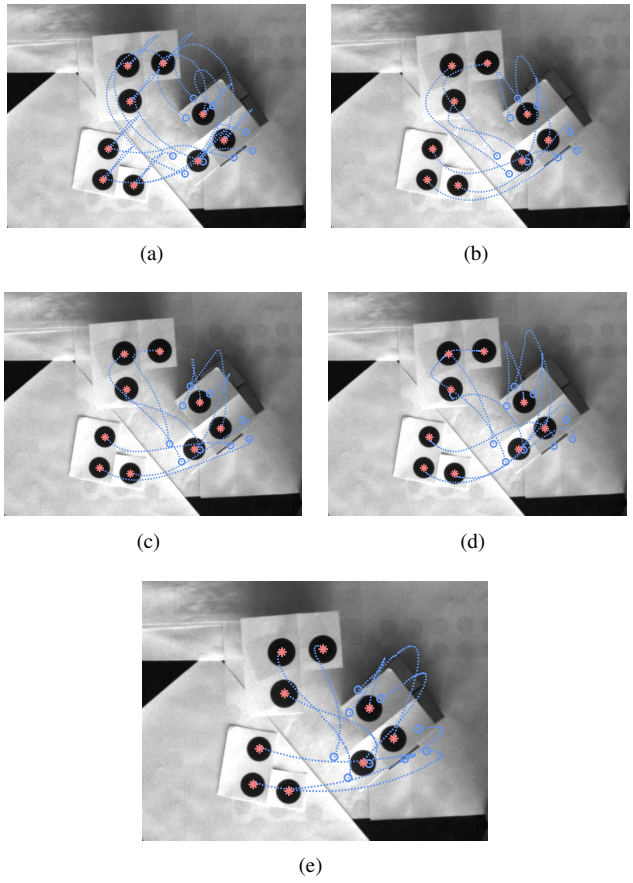

Figure 7. Experimental comparisons of the image trajectory for: (a) method 1, (b) method 2, (c) method 3, (d) method 4, (e) and proposed approach in the large rotation case.

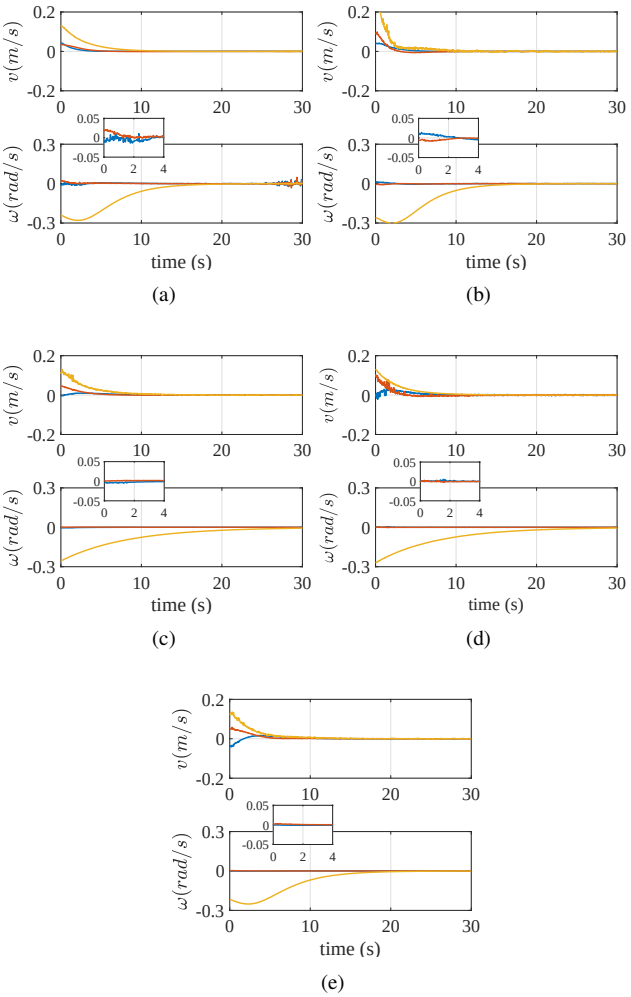

Figure 8. Experimental comparisons of the control inputs for: (a) method 1, (b) method 2, (c) method 3, (d) method 4, (e) and proposed approach in the large rotation case.

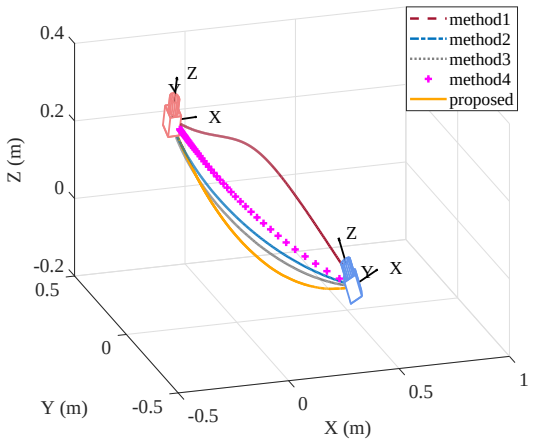

Figure 9. Cartesian space motion trajectories during the general motion case.

needs to be adjusted, but as presented in Figures 8(a) and 8(b) redundant rotation motion with respect to the other two axes exists in Method 1 and Method 2. As aforementioned, Method 1 and Method 2 define the visual features with homography elements and all the trifocal tensor elements, respectively. Since these features do not describe the pose information in a decoupling way, extra rotation along the x- and y-axes are introduced.

• *General Motion Case:* From Table 1, it is clear that both large translation and rotation errors along the three axes exist in this case. As illustrated in Figures 9 and 10, the Cartesian and image trajectories of Method 2, Method 3, Method 4, and proposed approach are well-behaved, while the trajectories

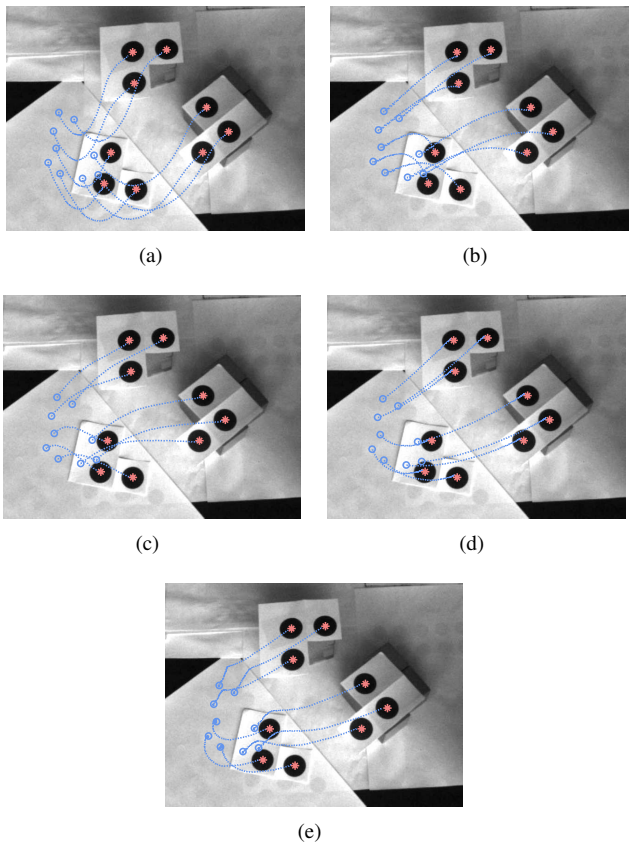

Figure 10. Experimental comparisons of the image trajectory for: (a) method 1, (b) method 2, (c) method 3, (d) method 4, (e) and proposed approach in the general motion case.

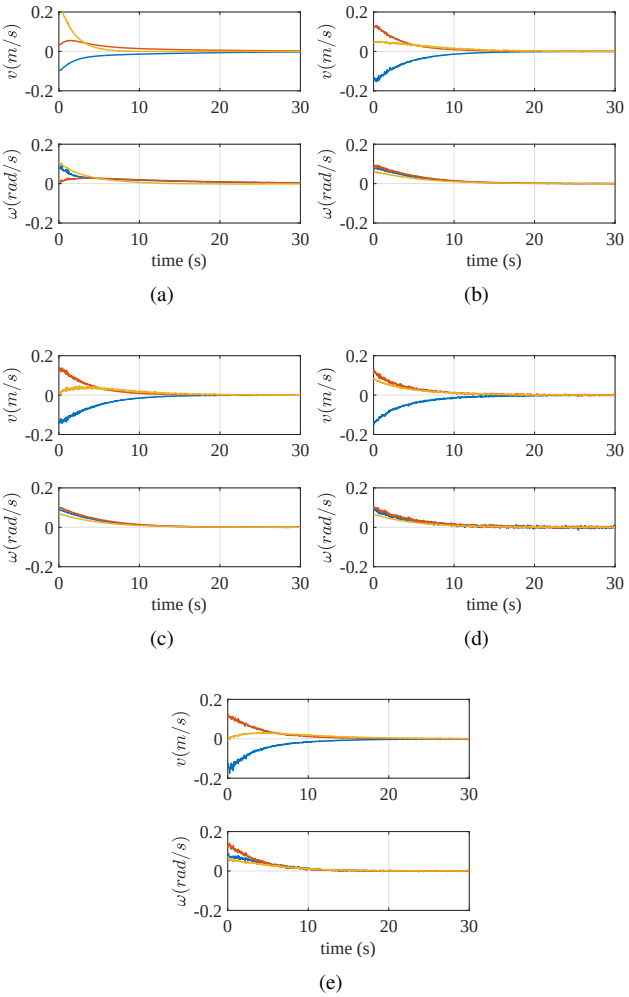

Figure 11. Experimental comparisons of the control inputs for: (a) method 1, (b) method 2, (c) method 3, (d) method 4, (e) and proposed approach in the general motion case.

corresponding to Method 1 contains overshooting. It is clear that for Method 1, some feature points move near to the image borders, which have the risk of leaving the camera view.

Although Method 1, Method 2, and the proposed approach all use elements extracted from homography or trifocal tensor to develop the control strategy, the performance of these three methods are quite different. Based on the above comparison results, it can be concluded that the visual features defined by the proposed approach show satisfactory decoupling characteristics. This is because the tensor elements used for the error system are selected by considering the geometric connotation, which guarantees that $e_{\mathcal{T}}(t)$ and $e_{\mathcal{Q}}(t)$ correspond to the translation and rotation errors, respectively. Owing to the decoupled system errors, the motion trajectory of the proposed approach is more efficient than the ones of Method 1 and Method 2 in aforementioned cases. Actually, the system errors constructed in Method 1 and Method 2 are coupled strongly, leading to the superfluous movement in specific configurations such as pure rotation and large translation cases.

6.2 Robustness analysis

6.2.1 Robustness to image noise. The proposed approach does not depend on SVD-based pose extraction during the control procedure, which can avoid complex on-line computation. In this subsection, the proposed approach is compared with SVD-based methods to evaluate the algorithm robustness in the presence of image noise. Since Method 3 and Method 4 rely on SVD, they are chosen to facilitate the comparison.

The implementation details are introduced now. A simulated free-floating camera controlled in 6 DOF is used for the regulation task. This camera is driven by these three methods, respectively, from a same initial pose to a desired pose. During the motion procedure, nine non-coplanar feature points are used to calculate the homography for Method 4 and the trifocal tensor for Method 3 and proposed approach. Gaussian noise with a standard deviation of 1 pixel and a mean of 0 pixel is added to the image coordinates of feature points. To compare the three methods rationally, all of them are executed by the same number of iterations, and the control gains are adjusted to ensure that the initial input amplitude of these three methods is similar. By tuning the iteration number and the gain parameters, it is also guaranteed that each method achieves stability (i.e., the system errors all converge to zero closely and remain in that situation for some time) before the end of iterations. Each method is tested by running 50 simulations, and the average results of the translation, rotation, and feature errors, i.e., e_T,

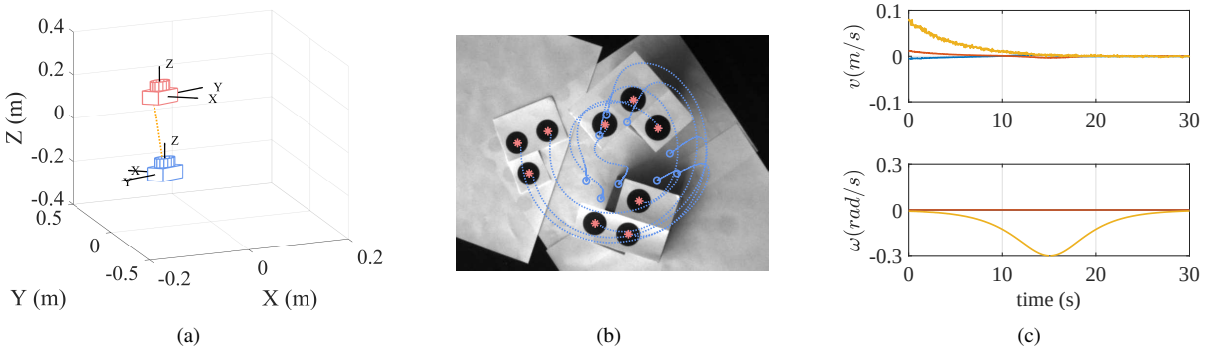

Figure 12. Results with initial system errors close to the equilirium Ω_2 and ${}^d t_i \neq 0_3$: (a) Camera motion in Cartesian space. (b) Image trajectory of the feature points. (c) Control inputs.

$e_R, e_I \in \mathbb{R}$, are calculated to facilitate the evaluation. e_T, e_R, and e_I are defined as

$$e_T \triangleq \|t_f - t_d\|$$

$$e_R \triangleq \|r_f - r_d\|$$

$$e_I \triangleq \frac{1}{9}\sum_{i=1}^{9} \|m_{fi} - m_{di}\|$$

where t_f, $t_d \in \mathbb{R}^3$ are the final and desired camera translational vectors, r_f, $r_d \in \mathbb{R}^3$ are the final and desired camera rotational vectors. t_f, t_d, r_f, and r_d are all expressed in the inertial coordinate frame. Additionally, m_{fi}, $m_{di} \in \mathbb{R}^3$ are the normalized Cartesian coordinates of the i-th feature point expressed in the final and desired camera frames, respectively.

Table 2. Comparison of different methods in the presence of image noise: e_T, e_R, and e_I represent the final translation, rotation, and feature errors, respectively.

	e_T	e_R	e_I
Method 3	0.0246	0.0647	0.0623
Method 4	0.0654	0.0573	0.0190
proposed approach	0.0255	0.0488	0.0333

The corresponding results are presented in Table 2. It can be seen that our approach and Method 3 present similar performance with respect to the final translation error, while for the rotation and feature errors, the proposed approach achieves better results than Method 3. Furthermore, by comparing the proposed approach with Method 4, it can be found that Method 4 achieves better convergence of the feature error, and the proposed approach shows smaller final translation and rotation errors. Method 4 directly uses one feature point as part of the system errors, and that is probably the reason why it presents nice performance with respect to feature error e_I.

The trifocal tensor $\mathcal{T}(t)$ is estimated by using the point sets $(m_c(t), m_d, m^*)$. As described in Section 3, the virtual normalized Cartesian coordinate m^* is calculated with the point correspondences m_i and m_d. Since m_i and m_d are extracted from the initial and desired images, the image noise has an indirect impact on m^*. To demonstrate the influence, simulation is conducted by using the "real" points. Precisely, the simulated camera is moved to the corresponding reference pose to capture the "real" points, and then the "real" points replace the virtual ones to facilitate the tensor computation. The average translation, rotation, and feature errors of 50 simulation runs are $[0.0252, 0.0379, 0.0227]$. Compared to the results shown in Table 2, it can be found that as expected, using the "real" points can achieve better control performance. This indicates that the virtual points are polluted by the image noise and are less precise than the "real" ones. Nonetheless, from a practical point of view, generating a virtual reference view to facilitate the tensor computation is much more convenient than truly capturing a third view.

6.2.2 Control performance near to unstable equilibria. As stated in Theorem 1, there exist three unstable equilibria of the closed-loop system (37). Since ${}^d R_*$ and ${}^d \bar{t}_*$ are selected as ${}^d R_* = {}^d R_i$ and ${}^d \bar{t}_* = -0.5\,{}^d \bar{t}_i$, it can be concluded from (58) that if the initial rotation errors are chosen as $[0°, 0°, 180°]$, or $[0°, 180°, 0°]$, or $[180°, 0°, 0°]$, then the initial system states are subject to the equilibrium Ω_2, or Ω_3, or Ω_4. To verify the effectiveness of proposed approach near to the unstable equilibria, an experiment with initial rotation errors $[0°, 0°, 178°]$ is conducted. Based on the results given in Figure 12, it can be found that due to the influence of Ω_2, the convergence rate at the beginning is quite slow, but as the camera pose moves away from Ω_2, the system errors get close to zero successfully. The obtained results essentially correspond to the theoretical analysis.

There exist translation errors between the initial and desired camera frames in the experiment shown in Figure 12. It is interesting to further evaluate the developed controller by selecting the initial pose errors as $[0m, 0m, 0m, 0°, 0°, 178°]$, which is very near to the Chaumette Conundrum (Corke and Hutchinson, 2001). Note that when the desired camera pose corresponds to a pure rotation (i.e., ${}^d t_i = 0_3$), the trifocal tensor $\mathcal{T}_i$ presented in (13) is equal to zero, and thus ${}^d R_i$ cannot be decomposed from $\mathcal{T}_i$. Under this circumstance, homography techniques are utilized to calculate ${}^d R_i$, and the rotation and translation between $\mathcal{F}_d$ and $\mathcal{F}^*$ are selected as ${}^d R_* = {}^d R_i$ and ${}^d t_* = [0.4m, -0.3m, -0.6m]^T$ to facilitate the tensor calculation. In addition, we find that Method 2 is also subject to the unstable equilibria of the proposed approach. Hence both Method 2 and the proposed approach are tested to

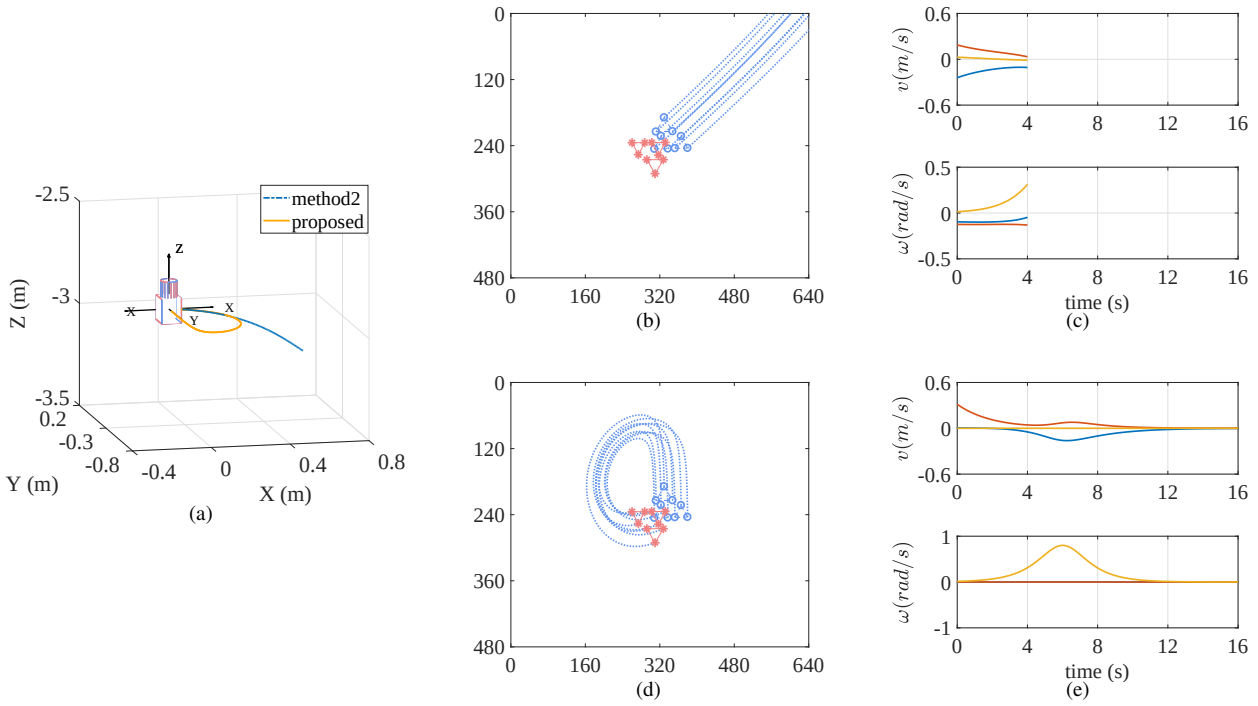

Figure 13. Simulation results with initial system errors close to the equilirium Ω_2 and ${}^{d}t_i = 0_3$: (a) Camera motion in Cartesian space. (b) Image trajectory of the feature points for method 2. (c) Control inputs for method 2. (d) Image trajectory of the feature points for proposed approach. (e) Control inputs for proposed approach.

better evaluate the control performance. The corresponding simulation results are shown in Figure 13. It can be seen that Method 2 produces both redundant translation and rotation motion, causing the feature points to leave the image. Although the proposed approach also introduces unnecessary translation motion, it can reduce the rotation errors along the z-axis effectively and drive the camera to the desired pose gradually. In fact, the performance of the proposed approach under this extreme case is not good enough, and this needs to be further studied.

By initializing the rotation errors close to $[0°, 180°, 0°]$ or $[180°, 0°, 0°]$, the target object will be out of the camera view, and thus the cases about the equilibria Ω_3 and Ω_4 are not further discussed.

6.2.3 Robustness to coarse ${}^{d}R_i$ and ${}^{d}\bar{t}_i$. In Algorithm 1, the pose signals ${}^{d}R_i$ and ${}^{d}\bar{t}_i$ need to be calculated to construct the reference frame $\mathcal{F}^*$. In order to test the robustness of the controller with respect to these two pose signals, supplementary errors are added to ${}^{d}R_i$ (5° on each axis) and ${}^{d}\bar{t}_i$ (5% on each axis). It is clear that coarse ${}^{d}R_i$ and ${}^{d}\bar{t}_i$ can reduce the estimation accuracy of tensor variables. Moreover, since ${}^{d}R_*$ is selected as ${}^{d}R_i$ in the experiment, the supplementary bias is also injected into ${}^{d}R_*$. The coarse tensor variables and ${}^{d}R_*$ will directly affect the control performance as described by (35). The experimental results are presented in Figure 14. Due to the introduction of pose errors, the camera motion path and the image trajectory of the feature points are different from the ones without supplementary errors. However, the camera still converges to the desired pose closely with the final positioning error being $[0.003m, -0.002m, 0.016m, -0.011°, -0.016°, 0.001°]$, which demonstrates the correct realization of the task.

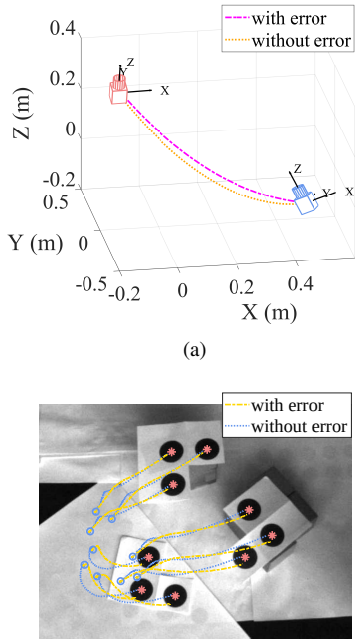

Figure 14. Results with supplementary errors added to ${}^{d}R_i$ and ${}^{d}\bar{t}_i$: (a) Camera motion in Cartesian space. (b) Image trajectory of the feature points.

6.2.4 Robustness to coarse intrinsic camera parameters. The algorithm robustness with respect to coarse intrinsic camera calibration matrix K is also considered. In this case, the image center is roughly used for the principal point, and

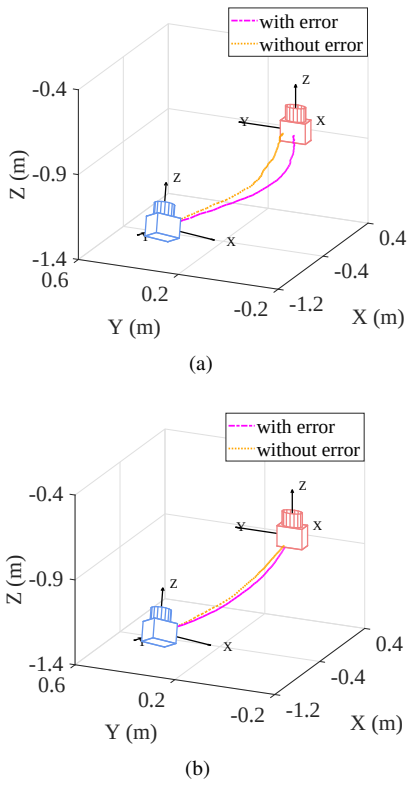

Figure 15. Results with supplementary errors added to intrisic camera calibration matrix K: (a) Camera motion in Cartesian space for Method 4. (b) Camera motion in Cartesian space for the proposed approach.

a supplementary error is added to the pixel and focal lengths (10%). It is obvious that the calibration accuracy of matrix K has a direct impact on the estimation of ${}^{d}R_i$, ${}^{d}\bar{t}_i$, tensor variables, etc. Therefore, the added error can trigger chain reactions that reduce the computation precision of related variables. To better illustrate the algorithm performance, Method 4, i.e, the 2.5 D visual servoing, and the proposed approach are tested under the same conditions. According to the results given in Figure 15, it can be found that for both methods, there exist differences between the camera trajectories resulting from the coarse and well calibrated K. Nonetheless, the deviation between the trajectories shown in Figure 15(b) is much smaller than the one presented in Figure 15(a), indicating that the proposed approach can achieve acceptable regulation result with coarse calibration matrix K.

6.3 Results of general scene

The last set of experiments are meant to illustrate the potential of our approach in addressing visual servoing under more realistic conditions compared to the use of blob features as done so far. To this end, we consider regulation of feature points lying on much less structured target objects.

In the first case, the observed scene consists of three drink boxes as shown in Figure 16(b). The contours of these boxes are tracked in real time with the aid of the moving-edges tracker available in ViSP, and then the intersection points of carton edges are extracted to calculate the trifocal tensor. The corresponding results are illustrated in Figure 16. Compared to the experiments based on blob features, it can be found that the control inputs presented in Figure 16(e) are not as smooth as the previous ones. That is because the scenario considered in this case contains more complex textured information, which influences the stability of feature extraction. From Figure 16(d), it can be seen that the system errors are close to zero after 15 s, implying satisfactory results can be achieved. The second case uses a soccer ball as the target object, which is presented in Figure 17(b). Extraction and tracking of the feature points are achieved by exploiting the Kanade-Lucas-Tomasi algorithm implemented in OpenCV. Figure 17 shows that both the Cartesian and image space errors converge, while the system errors and control inputs jitter near the end. Considering the less structured scenario, the jitter issue is caused by the accumulation errors of feature tracking. Actually, no all the detected feature points can be tracked consistently during the camera motion. As the tracking errors accumulate, the stability of trifocal tensor estimation reduces, leading to the jitter phenomenon near the end.

7 Conclusion

A 6 DOF visual servoing approach based on trifocal tensor was proposed in this paper. Considering the geometric connotation of trifocal tensor model, partial tensor elements were selected to define the visual features. Moreover, an adaptive controller was designed to accomplish the regulation task. The system stability was analyzed via the Lyapunov-based method. Existing controllers that use trifocal tensor (or homography) elements as feature information can only achieve local stability. However, the developed approach can ensure a larger convergence region theoretically, and the unstable equilibria of the error system were clearly derived. Several simulation and experimental scenarios were studied to qualitatively evaluate the performance of the proposed approach. Owing to the constructed visual feedback satisfied decoupling properties, it can be found from the first set of experiments that the motion trajectories of the developed approach were more efficient than the ones based on coupled visual features. The results also illustrated that the proposed approach was robust to image noise, coarse approximation of pose signals, and coarse intrinsic camera calibration matrix.

Nevertheless, the trifocal tensor based 6 DOF visual servoing still deserves further developments in some aspects. The trifocal tensor is calculated by using point correspondences in this paper, and thus the conventional feature detection and tracking are required during the control process. Obviously, the computation accuracy of trifocal tensor is sensitive to the feature detection and tracking errors. To enhance computation robustness, a possible way would be to estimate the trifocal tensor with dense information, such as image intensities (Benhimane and Malis, 2007; Silveira and Malis, 2012).

Although an adaptive update law is designed to compensate for the unknown distance, the estimated value is not convergent to the actual one. Note that if the unknown distance can be obtained accurately, then the structure of the observed scene can be recovered. To make the vision-based robotic system versatile, it would be interesting to

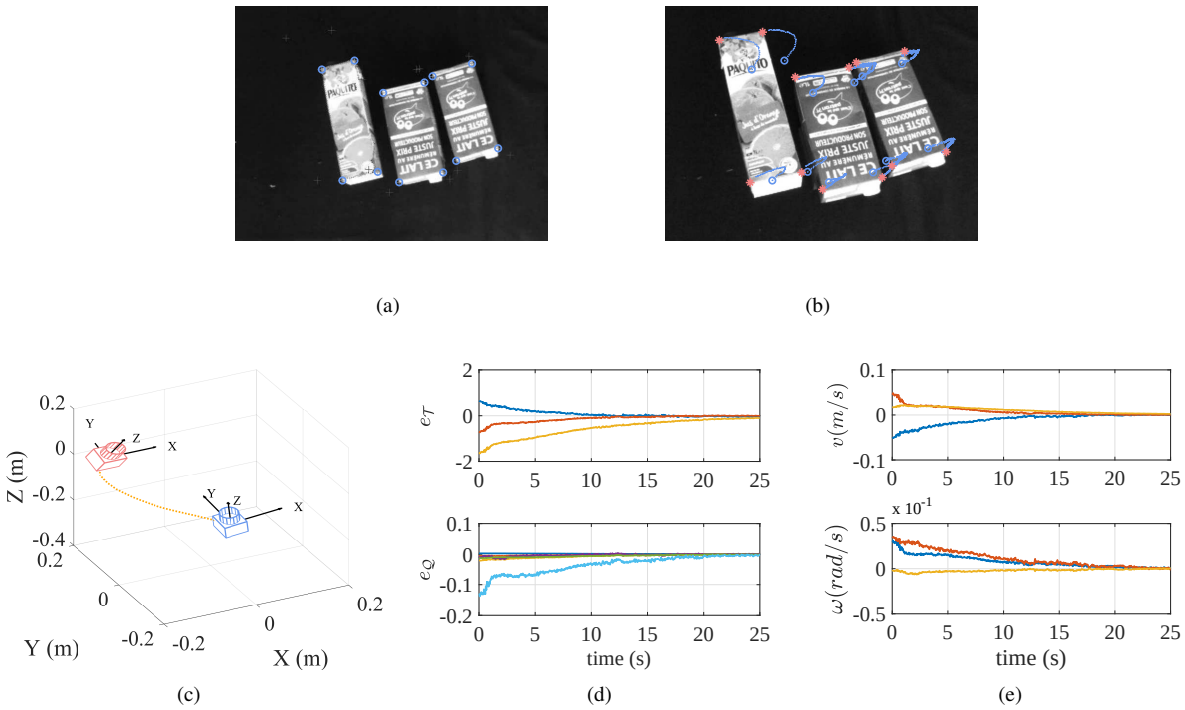

Figure 16. Regulation of feature points extracted from three drink boxes: (a) Initial image. (b) Image trajectory of the feature points. (c) Camera motion in Cartesian space. (d) System error convergence. (e) Control inputs.

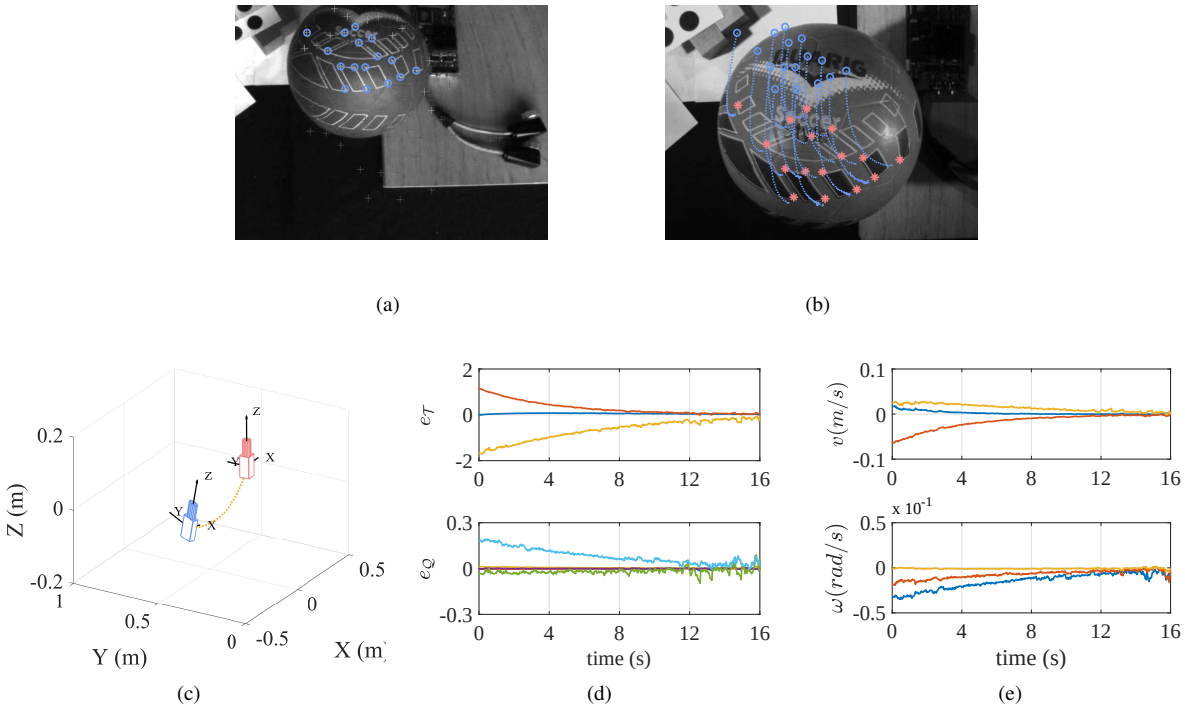

Figure 17. Regulation of feature points extracted from a soccer ball: (a) Initial image. (b) Image trajectory of the feature points. (c) Camera motion in Cartesian space. (d) System error convergence. (e) Control inputs.

simultaneously accomplish the distance estimation during the visual servoing. The active optimization strategy (Spica et al., 2017) or concurrent learning technique (Parikh et al., 2018) could be coupled into the proposed control scheme in this regard.

Funding

This work was supported in part by the National Natural Science Foundation of China under Grant 61433013, and in part by the scholarship from Zhejiang University.

References

Andreff N and Tamadazte B (2016) Laser steering using virtual trifocal visual servoing. *International Journal of Robotics Research* 35(6): 672–694.

Becerra HM, Hayet JB and Sagüés C (2014) A single visual-servo controller of mobile robots with super-twisting control. *Robotics and Autonomous Systems* 62(11): 1623–1635.

Becerra HM, López-Nicolás G and Sagüés C (2011) A sliding-mode-control law for mobile robots based on epipolar visual servoing from three views. *IEEE Transactions on Robotics* 27(1): 175–183.

Becerra HM and Sagüés C (2013) Exploiting the trifocal tensor in dynamic pose estimation for visual control. *IEEE Transactions on Control Systems Technology* 21(5): 1931–1939.

Benhimane S and Malis E (2007) Homography-based 2D visual tracking and servoing. *International Journal of Robotics Research* 26(7): 661–676.

Chaturvedi N, Sanyal AK, McClamroch NH et al. (2011) Rigid-body attitude control. *IEEE Control Systems Magazine* 31(3): 30–51.

Chaumette F and Hutchinson S (2006) Visual servo control part I: Basic approaches. *IEEE Robotics and Automation Magazine* 13(4): 82–90.

Chen J, Dawson DM, Dixon WE and Behal A (2005) Adaptive homography-based visual servo tracking for a fixed camera configuration with a camera-in-hand extension. *IEEE Transactions on Control Systems Technology* 13(5): 814–825.

Chen J, Jia B and Zhang K (2017) Trifocal tensor-based adaptive visual trajectory tracking control of mobile robots. *IEEE Transactions on Cybernetics* 47(11): 3784–3798.

Chen J, Jia B and Zhang K (2018) *Multi-View Geometry Based Visual Perception and Control of Robotic Systems*. Boca Raton: CRC Press.

Collewet C and Marchand E (2011) Photometric visual servoing. *IEEE Transactions on Robotics* 27(4): 828–834.

Corke PI and Hutchinson SA (2001) A new partitioned approach to image-based visual servo control. *IEEE Transactions on Robotics and Automation* 17(4): 507–515.

Dame A and Marchand E (2011) Mutual information-based visual servoing. *IEEE Transactions on Robotics* 27(5): 958–969.

Dixon WE, Dawson DM, Zergeroglu E and Behal A (2001) Adaptive tracking control of a wheeled mobile robot via an uncalibrated camera system. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics* 31(3): 341–352.

Fang Y, Dixon WE, Dawson DM and Chawda P (2005) Homography-based visual servo regulation of mobile robots. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics* 35(5): 1041–1050.

Fujita M, Kawai H and Spong MW (2007) Passivity-based dynamic visual feedback control for three-dimensional target tracking: Stability and L2-gain performance analysis. *IEEE Transactions on Control Systems Technology* 15(1): 40–52.

Hartley R (1997) Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision* 22(2): 125–140.

Hartley R and Zisserman A (2003) *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Horn RA and Johnson CR (2012) *Matrix Analysis*. Cambridge University Press.

Hutchinson S, Hager GD and Corke PI (1996) A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation* 12(5): 651–670.

Kallem V, Dewan M, Swensen JP, Hager GD and Cowan NJ (2007) Kernel-based visual servoing. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. San Diego, CA, USA, pp. 1975–1980.

Khalil HK (2002) *Nonlinear Systems*. Upper Saddle River, NJ: Prentice-Hall.

Lippiello V, Siciliano B and Villani L (2007) Position-based visual servoing in industrial multirobot cells using a hybrid camera configuration. *IEEE Transactions on Robotics* 23(1): 73–86.

Liu YH, Wang H, Wang C and Lam KK (2006) Uncalibrated visual servoing of robots using a depth-independent interaction matrix. *IEEE Transactions on Robotics* 22(4): 804–817.

López-Nicolás G, Guerrero JJ and Sagüés C (2010) Visual control through the trifocal tensor for nonholonomic robots. *Robotics and Autonomous Systems* 58(2): 216–226.

López-Nicolás G, Sagüés C and Guerrero JJ (2009) Parking with the essential matrix without short baseline degeneracies. In: *IEEE International Conference on Robotics and Automation*. Kobe, Japan, pp. 1098–1103.

Ma Y, Soatto S, Košecká J and Sastry SS (2003) *An Invitation to 3-D Vision: From Images to Geometric Models*. New York: Springer-Verlag.

Malis E and Chaumette F (2000) 2 1/2 D visual servoing with respect to unknown objects through a new estimation scheme of camera displacement. *International Journal of Computer Vision* 37(1): 79–97.

Malis E, Chaumette F and Boudet S (1999) 2-1/2-D visual servoing. *IEEE Transactions on Robotics and Automation* 15(2): 238–250.

Marchand E, Spindler F and Chaumette F (2005) ViSP for visual servoing: A generic software platform with a wide class of robot control skills. *IEEE Robotics and Automation Magazine* 12(4): 40–52.

Mariottini GL, Oriolo G and Prattichizzo D (2007) Image-based visual servoing for nonholonomic mobile robots using epipolar geometry. *IEEE Transactions on Robotics* 23(1): 87–100.

Parikh A, Kamalapurkar R and Dixon WE (2018) Target tracking in the presence of intermittent measurements via motion model learning. *IEEE Transactions on Robotics* 34(3): 805–819.

Sabatta D and Siegwart R (2013) Vision-based path following using the 1D trifocal tensor. In: *IEEE International Conference on Robotics and Automation*. Karlsruhe, Germany, pp. 3095–3102.

Shademan A and Jägersand M (2010) Three-view uncalibrated visual servoing. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Taipei, Taiwan, pp. 6234–6239.

Silveira G and Malis E (2012) Direct visual servoing: Vision-based estimation and control using only nonmetric information. *IEEE Transactions on Robotics* 28(4): 974–980.

Slotine JJE and Li W (1991) *Applied nonlinear control*. Englewood Cliffs, NJ: Prentice-Hall.

Spica R, Giordano PR and Chaumette F (2017) Coupling active depth estimation and visual servoing via a large projection operator. *International Journal of Robotics Research* 36(11): 1177–1194.

Tayebi A, Roberts A and Benallegue A (2013) Inertial vector measurements based velocity-free attitude stabilization. *IEEE Transactions on Automatic Control* 58(11): 2893–2898.

Wilson WJ, Hulls CCW and Bell GS (1996) Relative end-effector control using Cartesian position based visual servoing. *IEEE Transactions on Robotics and Automation* 12(5): 684–696.

Zhang K, Chen J and Chaumette F (2018) Visual servoing with trifocal tensor. In: *IEEE Conference on Decision and Control*. Miami Beach, FL, USA.

Zhang K, Chen J and Jia B (2017) Asymptotic moving object tracking with trajectory tracking extension: A homography-based approach. *International Journal of Robust and Nonlinear Control* 27(18): 4664–4685.

Appendix A. Potential degeneration of trifocal tensor

In this paper, the trifocal tensor is used to encapsulate the geometric relationship across three views and is estimated via point correspondences based on the conventional model presented in (3). However, from Figure 1, if two of the three frames $\mathcal{F}_c$, $\mathcal{F}_d$, and $\mathcal{F}^*$ are same, there may exist degeneration for the tensor estimation. More precisely, in the case that $\mathcal{F}^*$ and $\mathcal{F}_d$ are coincident, it can be found from Figure 1 that ${}^dR_* = I_{3\times3}$, ${}^dt_* = 0_3$, and $m^* = m_d$. Then, (1) can be rewritten as

$$\mathcal{T}_{(j)} = -{}^ct_* I^T_{3\times3(j)}, \tag{59}$$

indicating that the trifocal tensor $\mathcal{T}(t)$ only contains the translation information ${}^ct_*(t)$. Based on (59), (3) can be simplified as

$$[m_c]_\times \left(-{}^ct_* m^{*T}\right)[m_d]_\times = 0_{3\times3}. \tag{60}$$

Since $m^{*T}[m_d]_\times = m_d^T[m_d]_\times = 0_{1\times3}$, ${}^ct_*(t)$ cannot be estimated from (60), i.e., $\mathcal{T}(t)$ cannot be estimated correctly from the model (3). Without loss of generality, the case that $\mathcal{F}^*$ and $\mathcal{F}_c$ are coincident can be analyzed by the same way.

Note that if $\mathcal{F}_c$ coincides with $\mathcal{F}_d$, the model (3) is convenient for estimating the trifocal tensor. Specifically, under this circumstance, ${}^cR_* = {}^dR_*$, ${}^ct_* = {}^dt_*$, and $m_c = m_d$. The trifocal tensor given in (1) can be reformulated as

$$\mathcal{T}_{(j)} = {}^dR_{*(j)}\,{}^dt_*^T - {}^dt_*\,{}^dR^T_{*(j)} = [x_j]_\times \tag{61}$$

where $x_j \in \mathbb{R}^3$ is given by

$$x_j \triangleq \begin{bmatrix} {}^dR_{*(3j)}\,{}^dt_{*(2)} - {}^dR_{*(2j)}\,{}^dt_{*(3)} \\ {}^dR_{*(1j)}\,{}^dt_{*(3)} - {}^dR_{*(3j)}\,{}^dt_{*(1)} \\ {}^dR_{*(2j)}\,{}^dt_{*(1)} - {}^dR_{*(1j)}\,{}^dt_{*(2)} \end{bmatrix}. \tag{62}$$

Using (61), (3) can be rewritten as

$$[m_c]_\times \left(\sum_{j=1}^{3} m^*_{(j)}[x_j]_\times\right)[m_d]_\times = 0_{3\times3}. \tag{63}$$

Based on $m_c = m_d$, (63) can be transformed into the following linear expression:

$$\underbrace{\begin{bmatrix} m^*_{(1)}m_d m_d^T & m^*_{(2)}m_d m_d^T & m^*_{(3)}m_d m_d^T \end{bmatrix}}_{B} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0_3. \tag{64}$$

It can be found from (64) that $B \in \mathbb{R}^{3\times9}$ has rank one, i.e., each pair of point correspondences provides one linearly independent equation to estimate x_1, x_2, and x_3. Precisely, the linear independent equation extracted from (64) is given by

$$\begin{bmatrix} m^*_{(1)}m_d^T & m^*_{(2)}m_d^T & m^*_{(3)}m_d^T \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0. \tag{65}$$

Provided enough pairs of point correspondences (at least 8), a homogeneous set of equations as (65) can be obtained, and then the trifocal tensor can be computed effectively up to a scale. Thus, the model (3) is applicable to tensor estimation in the case that $\mathcal{F}_c$ coincides with $\mathcal{F}_d$.

Appendix B. Proof of Property 1

Proof. According to (26) and (34), $L_\mathcal{Q}(t)$ can be rewritten as

$$L_\mathcal{Q} = \frac{1}{d^*}\begin{bmatrix} {}^dt_*^T\left[{}^dR_{*(1)}\right]_{\times(1)}\left[{}^cR_{*(1)}\right]_\times \\ {}^dt_*^T\left[{}^dR_{*(2)}\right]_{\times(1)}\left[{}^cR_{*(2)}\right]_\times \end{bmatrix}. \tag{66}$$

Then, it can be derived that

$$\begin{aligned} L_\mathcal{Q}^T L_\mathcal{Q} = \frac{1}{d^{*2}}\Bigg(&\left({}^dt_*^T\left[{}^dR_{*(1)}\right]_{\times(1)}\right)^2\left[{}^cR_{*(1)}\right]_\times^T\left[{}^cR_{*(1)}\right]_\times \\ &+\left({}^dt_*^T\left[{}^dR_{*(2)}\right]_{\times(1)}\right)^2\left[{}^cR_{*(2)}\right]_\times^T\left[{}^cR_{*(2)}\right]_\times\Bigg). \end{aligned} \tag{67}$$

For any $y \in \mathbb{R}^3$, it is clear that

$$\begin{aligned} &y^T L_\mathcal{Q}^T L_\mathcal{Q} y \\ &= \frac{1}{d^{*2}}\Bigg(\left({}^dt_*^T\left[{}^dR_{*(1)}\right]_{\times(1)}\right)^2 y^T\left[{}^cR_{*(1)}\right]_\times^T\left[{}^cR_{*(1)}\right]_\times y \\ &+\left({}^dt_*^T\left[{}^dR_{*(2)}\right]_{\times(1)}\right)^2 y^T\left[{}^cR_{*(2)}\right]_\times^T\left[{}^cR_{*(2)}\right]_\times y\Bigg). \end{aligned} \tag{68}$$

As stated in Section 4.3, ${}^dt_*^T\left[{}^dR_{*(1)}\right]_{\times(1)} \neq 0$ and ${}^dt_*^T\left[{}^dR_{*(2)}\right]_{\times(1)} \neq 0$. Additionally, we have

$$\begin{aligned} y^T\left[{}^cR_{*(1)}\right]_\times^T\left[{}^cR_{*(1)}\right]_\times y \geq 0 \\ y^T\left[{}^cR_{*(2)}\right]_\times^T\left[{}^cR_{*(2)}\right]_\times y \geq 0. \end{aligned} \tag{69}$$

For $y \neq 0_3$, $y^T\left[{}^cR_{*(1)}(t)\right]_\times^T\left[{}^cR_{*(1)}(t)\right]_\times y = 0$ if and only if y and ${}^cR_{*(1)}(t)$ are collinear. Similarly, $y^T\left[{}^cR_{*(2)}(t)\right]_\times^T\left[{}^cR_{*(2)}(t)\right]_\times y = 0$ is satisfied, for some $y \neq 0_3$, if and only if y and ${}^cR_{*(2)}(t))$ are collinear. Since ${}^cR_{*(1)}(t)$ and ${}^cR_{*(2)}(t)$ are linearly independent, they cannot be both collinear to the same y. Based on (68), (69), and the above analysis, it can be concluded that for $y \neq 0_3$, $y^T L_\mathcal{Q}^T(t) L_\mathcal{Q}(t) y > 0$, i.e., $L_\mathcal{Q}^T(t)L_\mathcal{Q}(t)$ is symmetric and positive definite. Therefore, the corresponding inverse matrix $\left(L_\mathcal{Q}^T(t)L_\mathcal{Q}(t)\right)^{-1}$ is also symmetric and positive definite (Horn and Johnson, 2012). □